From Financial Reports to Macroeconomic Forecasts: An LLM-Based Approach

Lusha Xu *

October 30, 2026

Abstract

Do qualitative narratives in corporate financial reports carry forward-looking information about the aggregate economy? Using a large language model and a structured questionnaire, I extract 38 filing-based measures from 60,236 Form 10-K and 10-Q filings submitted by S&P 500 firms between 1996 and 2026. Each measure is aligned by its public EDGAR filing date and distinguishes between explicitly stated disclosures and contextual model inferences. This corporate soft information is economically meaningful at both the micro and macro levels: the extracted measures predict subsequent firm-level outcomes, and their aggregates closely track target macroeconomic series. Using variables selected through 2012, I construct weekly leading and coincident indices to evaluate real-time recursive forecasts. Over 2013–2019, incorporating these indices reduces mean squared forecast error relative to an autoregressive benchmark by 17 percent for the current quarter and 18 percent one quarter ahead, averaged over weeks 5–13. The indices also improve GDP quantile forecasts, with larger gains in lower-tail accuracy. Beyond GDP, filing-based measures improve industrial production point forecasts and year-ahead core inflation quantile forecasts. Labor-market gains are less consistent, and production and inflation point-forecast gains weaken after 2020. These findings show that corporate soft information complements aggregate statistics by capturing firms' assessments of business conditions and anticipated adjustments in economic activity.

Keywords: Large Language Models, Macroeconomic Forecasting, Factor Models, Corporate Disclosures, Narrative Economics.
JEL Classification: E37, C55, G17, C53, M41.

*Department of Economics, University of California, Riverside. Email: lxu099@ucr.edu

1 Introduction

Official macroeconomic statistics measure aggregate activity with a substantial lag: the Bureau of Economic Analysis (BEA) releases its advance GDP estimate roughly a month after quarter-end. A large literature forecasts and nowcasts these series using higher-frequency hard data, typically by extracting common factors from macroeconomic and financial panels (Stock and Watson, 2003, Giannone et al., 2008, Bańbura et al., 2013). Yet this paradigm relies almost exclusively on realized quantities, overlooking some of the most forward-looking information in the economy: the qualitative disclosures of corporate managers. When executives describe softening demand, tightening balance sheets, or supply bottlenecks in their regulatory filings, they summarize private, granular knowledge of their operating environment before it appears in aggregate statistics. Historically, extracting consistent measures from tens of thousands of unstructured filings has been impractical.

This paper shows that the soft information embedded in routine corporate financial reports contains real-time predictive power for macroeconomic activity and introduces a transparent, scalable framework to unlock it. Using a comprehensive 50-item questionnaire and a structured prompting protocol, I direct a Large Language Model (LLM) to read full-length Form 10-K and 10-Q filings. The sample covers every filing by an S&P 500 member between January 1996 and September 2026—60,236 filings across 1,159 firms. From each filing, the model extracts 38 quantitative economic variables capturing guidance, operating plans, supply chains, risks, policy exposure, and sentiment. Dating each observation strictly by its public EDGAR filing date yields an auditable panel of roughly three million structured answers.

Why does soft information from corporate filings add value beyond realized accounting numbers? Whereas financial statements record historical outcomes, narrative disclosures reveal the *mechanisms* driving those outcomes and management's expectations for the future. For instance, a rise in book inventories on a balance sheet is economically ambiguous: it could reflect optimistic stock-building ahead of strong sales, an unexpected collapse in final demand, or unfinished goods trapped by supply-chain frictions. The accompanying narrative resolves this ambiguity by simultaneously capturing customer demand, inventory adequacy, pricing power, and production plans.

Before forecasting, I validate the LLM-extracted measures against firm and aggregate conditions. The measures predict subsequent firm outcomes. In regressions with firm and calendar-quarter fixed effects and controls for recent sales growth, 28 of the 38 measures significantly predict their assigned outcome over the following year. Hiring intentions predict employment growth, pricing power predicts changes in operating margins, and inventory plans predict inventory growth. Twenty-two of these relationships remain significant using inferred answers alone. The aggregates also track related macroeconomic benchmarks: production plans correlate with industrial-production growth at 0.78, hiring intentions with payroll growth at 0.73, and inventory plans with inventory growth at 0.86.

I next evaluate whether this firm-level soft information improves real-time macroeconomic nowcasts and forecasts. For GDP, I construct weekly leading and coincident indices, selecting variables on 1996–2012 and holding them fixed thereafter. The leading measures consist mainly of changes in

revenue guidance, inventories relative to demand, wage expectations, and hiring intentions, together with supply-chain risk. The coincident measures describe production, investment, and uncertainty. Separate dynamic factor models accommodate the uneven arrival of filings (Bańbura and Modugno, 2014). Forecasts are estimated recursively using information available at each forecast date.

During 2013–2019, the filing measures reduce the mean squared forecast error (MSFE) of a real-time autoregressive benchmark by 17% for the current quarter (nowcast) and 18% one quarter ahead once the prior quarter's advance GDP estimate is published, and by 10–22% in earlier weeks of the quarter. The two indices play distinct, complementary roles: the coincident flow factor drives the current-quarter nowcast as new filings arrive each week, while the leading stock factor drives the one-quarter-ahead forecast. Furthermore, incorporating the filing indices into quantile regressions (Adrian et al., 2019) improves the entire predictive distribution of GDP growth by 18–28% and reduces downside-weighted (left-tail) forecast loss by 30–36%, demonstrating that corporate narratives are particularly informative about macroeconomic tail risks.

Extending the framework to monthly macroeconomic targets reveals where corporate narratives are most informative. Because S&P 500 disclosures heavily reflect goods production, capital investment, and pricing decisions, variable-specific MIDAS regressions on the weekly filing series significantly improve point forecasts of industrial production at all horizons (reducing MSFE by 11–29% in 2013–2019 and 37–57% in a 2009–2012 validation sample), while two-index quantile regressions significantly improve twelve-month-ahead density forecasts of core CPI inflation. By contrast, the filings do not consistently improve out-of-sample forecasts of nonfarm payroll employment, reflecting the fact that large public firms can adjust hours and utilization before headcount and represent only one segment of aggregate labor flows. Across targets, the out-of-sample forecasting gains largely attenuate after 2020, when pandemic shutdowns, unprecedented fiscal transfers, and severe supply-chain swings altered historical lead-lag relationships.

Finally, robustness and look-ahead audits confirm the reliability of the empirical design. A truncation test removing later-dated inputs in seventeen historical target–date cases reproduces forecasts to numerical tolerance (10^{-9} to 10^{-14}), while re-estimating using only explicitly stated disclosures preserves the GDP forecasting gains. Furthermore, pre-specified filing composites predict Survey of Professional Forecasters (SPF) errors and retain predictive power for output, investment, employment, and core inflation over standard macroeconomic controls.

This paper makes one methodological and two empirical contributions to narrative economics and forecasting. Methodologically, I use generative LLMs to move beyond scalar sentiment and topic models (Loughran and McDonald, 2011, Baker et al., 2016, Hassan et al., 2019), extracting an auditable panel of structured economic indicators from full-length regulatory filings. Empirically, I show that properly aligned corporate soft information captures firm fundamentals and aggregate dynamics, providing incremental forecasting power for output, industrial production, and inflation risk.

2 Literature Review

This paper contributes to three strands of literature. The first links firm-level financial reporting to macroeconomic forecasting. Seminal work demonstrates that aggregate accounting metrics—such as earnings growth and dispersion—predict future GDP, unemployment, and producer price shocks (Konchitchki and Patatoukas (2014), Nallareddy and Ogneva (2017), Shivakumar and Urcan (2017)). Recently, this literature has shifted toward high-frequency integration, utilizing frameworks like the Dynamic Factor Model (DFM) to show that weekly or daily accounting signals significantly improve macroeconomic nowcasts and forecasts (Abdalla et al. (2021), Pettenuzzo et al. (2024)). However, these studies rely exclusively on numerical financial data, overlooking the rich, forward-looking narratives embedded in corporate reports. This leaves a critical gap to explore textual disclosures as an alternative and potential predictive data source for anticipating macroeconomic shifts.

The second strand relates to textual analysis in economics and finance, which has evolved from dictionary-based word frequencies to advanced machine learning algorithms. Early applications relied on predefined or custom dictionaries to measure sentiment, political risk, and economic uncertainty (Tetlock (2007), Loughran and McDonald (2011), Baker et al. (2016), Ahir et al. (2022), Caldara and Iacoviello (2022), Hassan et al. (2019)). As computational capabilities advanced, researchers moved beyond word counts toward unsupervised topic modeling. Building on the Latent Dirichlet Allocation (LDA) model of Blei et al. (2003), subsequent studies applied LDA to separate economic shocks in FOMC statements Hansen and McMahon (2016) and track macroeconomic trends using expansive news archives Bybee et al. (2024).

The third strand applies Large Language Models (LLMs) to economics and finance (Vaswani et al. (2017), Devlin et al. (2019), Araci (2019)). Across these disciplines, the literature generally bifurcates into two methodological approaches. The first deploys LLMs directly as end-to-end forecasting engines or proxy simulation agents. For example, Carriero et al. (2024) utilize LLMs directly as numerical time-series algorithms, feeding them historical macroeconomic data (e.g., FRED-MD) to predict future values. Taking a behavioral approach, Faria-e Castro and Leibovici (2024), Zarifhonarvar (2025), and Hansen et al. (2026) prompt LLMs to simulate survey respondents and directly forecast inflation. Because final predictions rely on opaque neural network weights, these act as econometric "black boxes." To maintain transparency, the second approach uses LLMs strictly for feature extraction—translating unstructured text into structured variables evaluated within traditional econometric frameworks. For instance, in asset pricing, Lopez-Lira and Tang (2023) use ChatGPT to extract sentiment scores from financial headlines to predict stock returns. In macroeconomics, this text-to-data approach has been successfully leveraged to extract economic beliefs from news Bybee (2023), decipher central bank policy stances Hansen and Kazinnik (2024), construct newspaper-based uncertainty indices Audrino et al. (2024), and quantify geoeconomic pressure Clayton et al. (2025).

My paper is most closely related to Phan (2025), who forecasts macroeconomic activity using accounting-topic attention and sentiment, and Isler (2025), who forecasts GDP downside risk using dictionary-based tone changes. I complement these studies by extracting 38 structured measures of

firms' economic conditions and plans, distinguishing demand, costs, production, and hiring rather than summarizing them through tone or topic prevalence. Each extraction includes supporting evidence and distinguishes explicit disclosures from contextual inference. After validating these measures against firm outcomes and aggregate benchmarks, I assess their contribution to point and distributional forecasts of first-release output, inflation, and labor-market outcomes.

3 Extracting Soft Information from Financial Reports

3.1 Data

Corporate Disclosures The primary textual dataset consists of 10-K and 10-Q filings from January 1996 to September 2026, obtained from the U.S. Securities and Exchange Commission (SEC) EDGAR database. U.S. public companies are required to file these reports, which contain financial statements, accompanying notes, and management's discussion and analysis (MD&A) of the firm's financial condition and operating performance (U.S. Securities and Exchange Commission, 2021). Form 10-K provides a comprehensive annual disclosure, including audited financial statements, whereas Form 10-Q provides an unaudited quarterly update on the firm's financial condition and operations.

The sample consists of every filing made by a firm while it was a member of the S&P 500: 60,236 filings by 1,159 firms. Membership is determined at the filing date from the weekly index constituent lists. Firms enter the sample only while they are S&P 500 members, including those that subsequently exit the index. Large firms matter for aggregate fluctuations: Gabaix (2011) finds that idiosyncratic movements among the largest 100 U.S. firms explain approximately one-third of the variation in aggregate output and the Solow residual, and the S&P 500 accounts for roughly 80% of total U.S. equity market capitalization (S&P Dow Jones Indices, 2026). The sample covers all major sectors of the economy (Appendix A).

The standardized structure and forward-looking content of 10-K and 10-Q filings make them well suited for constructing systematic measures of firm-level economic information. Compared with earnings calls, which are relatively conversational and unstructured, these filings follow standardized disclosure requirements that facilitate systematic LLM-based information extraction across firms and over time. The filings contain information on realized performance, operating conditions, business risks, economic exposures, sentiment, and management's expectations about future business conditions and industry prospects. Aggregating these measures across large firms may therefore provide a useful signal about broader economic conditions. For example, Amazon's 2024 10-K states in its MD&A section:

> This Annual Report on Form 10-K includes *forward-looking statements* within the meaning of the Private Securities Litigation Reform Act of 1995. All statements other than statements of historical fact, *including statements regarding guidance, industry prospects, or future results of operations or financial position,* made in this Annual Report on Form

> 10-K are forward-looking. We use words such as *anticipates, believes, expects, future, intends, and similar expressions to identify forward-looking statements* [...] Actual results and outcomes could differ materially for a variety of reasons, including [...] changes in global economic conditions and customer demand and spending, inflation, interest rates, regional labor market constraints, [and] world events [...]

Aggregating these forward-looking assessments across large firms may therefore provide a useful, systematic signal about broader economic conditions.

Macroeconomic Data and Temporal Alignment

To measure aggregate economic activity, I use real-time data vintages from the Federal Reserve Bank of St. Louis's ALFRED database: real GDP at quarterly frequency, and industrial production, core CPI, nonfarm payroll employment and the unemployment rate at monthly frequency. The forecasting targets are year-over-year growth rates as first published (for the unemployment rate, the 12-month change in percentage points).

The Bureau of Economic Analysis (BEA) releases GDP estimates in sequential stages—the advance estimate of 2025:Q4 GDP in late January 2026, followed by second and third estimates in February and March—the data are subject to substantial subsequent revisions. In the nowcasting model, I utilize real-time data vintages.

Furthermore, I strictly align the corporate textual data based on the *filing date* rather than the reporting period. Because a Q1 10-Q is typically filed in early May (40 to 45 days post-quarter-end), it is structurally unavailable to forecast Q1 advance GDP released in late April. Therefore, filings are mapped to the macro information set based exclusively on the date they become publicly available on EDGAR, ensuring the model generates forecasts using only the exact information set available to economic agents in real time.

3.2 Methodology

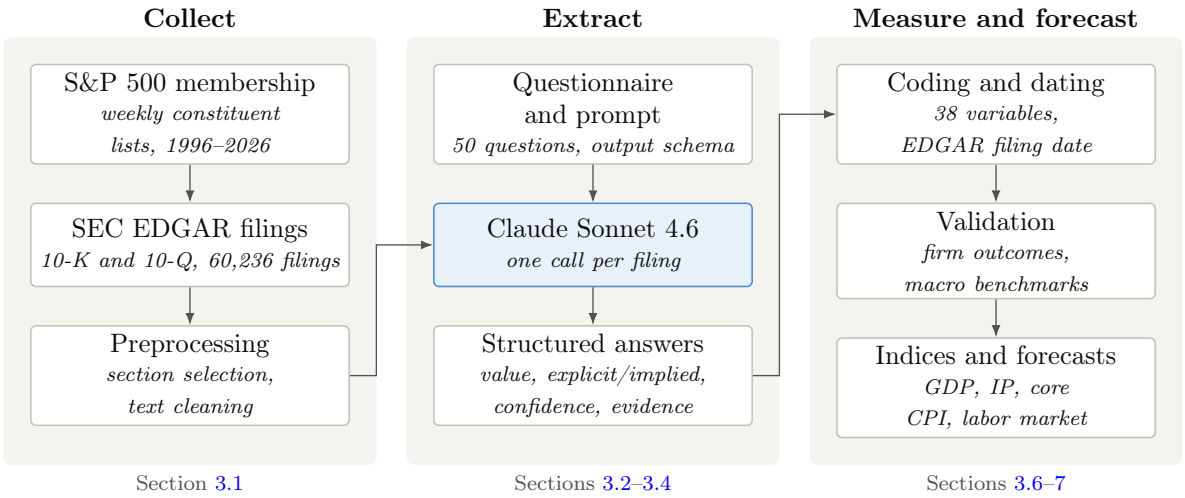
To construct the firm-level panel dataset, I employ a generative large language model to extract structured information from corporate disclosures. Traditional natural language processing tools, such as bag-of-words methods, rely on isolated term frequencies and fail to resolve semantic ambiguities—for example, distinguishing a financial "bank" from a river "bank." Transformer-based models resolve this through context-aware self-attention. Yet encoder-only architectures such as BERT and FinBERT remain unsuitable for my task: they are constrained by a strict 512-token[1] limit, rendering them impractical for multi-page disclosures, and require resource-intensive fine-tuning on hand-labeled data. By contrast, modern generative LLMs process far larger context windows and perform zero-shot[2] extraction of heterogeneous categorical and numerical variables.

[1] Transformer models split text into words or subwords, termed *tokens*, which are then mapped to numerical vectors that the model processes.

[2] The term *zero-shot* indicates that the model performs the extraction without task-specific fine-tuning or a labeled training corpus; each task is specified through natural-language instructions in the prompt. This contrasts with supervised approaches such as fine-tuned BERT/FinBERT, which require hand-labeled training data for each task.

Such models have been shown to exceed encoder-based architectures such as BERT or FinBERT on economic text-interpretation tasks (Hansen and Kazinnik, 2024, Lopez-Lira and Tang, 2023).

Figure 1: Data construction and LLM-based extraction procedure



I use Claude Sonnet 4.6 (Anthropic, 2026), accessed through the API, to extract information from SEC filings. A primary empirical challenge in the textual analysis of corporate disclosures is document length; 10-Q filings frequently span 50,000 tokens, and 10-K filings routinely exceed 100,000 tokens. Claude 4.6 Sonnet features a 200,000-token context window, I can ingest the entirety of each filing without relying on ad-hoc text chunking or truncation algorithms. This full-context processing is critical, as accurately capturing my target variables requires the model to synthesize information dispersed across multiple sections of the same document. While frontier models such as Claude 4 Opus offer marginal performance gains on highly complex reasoning tasks, their unit economics (costing approximately five times more per token) render them cost-prohibitive for a corpus of 60,236 filings. [3]

Figure 1 summarizes the integrated data acquisition and structured extraction pipeline. The workflow has three stages: collecting the filings, extracting structured answers from them, and turning the answers into measures and forecasts. In the first stage, I identify the filings of every S&P 500 member and preprocess the raw EDGAR files to filter the relevant textual sections and strip extraneous formatting artifacts (see Appendix B for exact cleaning procedures). Condensing the documents serves a dual purpose: it provides the model with a cleaner context window—mitigating noise-induced extraction errors—and substantially reduces token expenditures. For the second stage,

The prompt includes one worked example for a single variable to fix the output format, but provides no labeled training examples for the extraction tasks themselves.

[3] In pilot tests on a subset of filings, I also evaluated alternative models. While OpenAI's GPT-4o and DeepSeek-V3 performed adequately on simple extraction tasks, they exhibited lower reliability in adhering to the strict output schema required for automated parsing. Smaller, cost-optimized models such as Claude Haiku and GPT-4o Mini proved unviable for this task, generating unacceptably high rates of omitted values and formatting drift.

I design a 50-variable structured questionnaire and prompt that defines the target macroeconomic concepts, specifies formatting constraints, and establishes rules for handling missing values (detailed in Sections 3.3 and 3.4). In the final extraction step, Claude 4.6 Sonnet processes each cleaned filing alongside the prompt, outputting a standardized JSON object containing the target variables, supporting textual evidence, and model confidence scores. The third stage, in which the answers are coded, dated, validated and aggregated into the indices used for forecasting, is described in Sections 3.6–7.

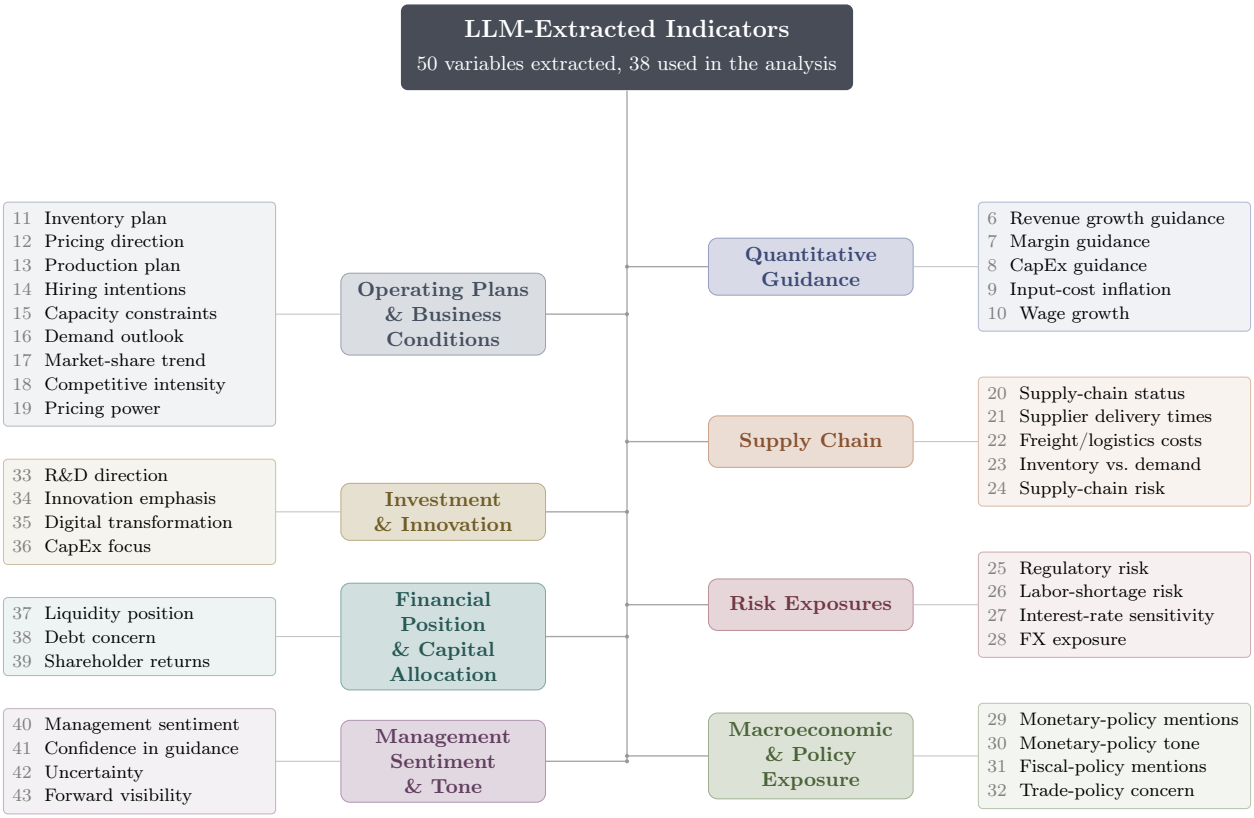
3.3 Questionnaire Design

The questionnaire is designed to specify the information and variables that the LLM extracts from each filing (see Appendix C for the complete questionnaire). I partition these variables into three functional groups. Questions Q1–Q5 capture firm identifiers and reporting periods used to construct the firm-period panel. Questions Q6–Q43 constitute the core empirical dataset: 38 substantive economic variables categorized into eight thematic groups, mapped in Figure 2. These groups span (1) Quantitative Guidance, (2) Operating Plans & Business Conditions, (3) Supply Chain, (4) Risk Exposures, (5) Macroeconomic & Policy Exposure, (6) Investment & Innovation, (7) Financial Position & Capital Allocation, and (8) Management Sentiment & Tone. For the remaining variables, Questions Q44–Q47 collect supplemental free-text information retained for validation and manual audit, while Q48–Q50 record geographic exposure and sector classifications.

The selection of the 38 core variables is guided by two primary criteria. First, they are firmly grounded in SEC disclosure mandates, making systematic extraction feasible across a broad set of filings; Appendix E.1 documents the regulatory basis and expected filing location for each variable. Second, I draw on prior work that uses corporate text to construct related firm-level measures, such as management sentiment and input cost inflation (Windsor and Zang, 2023, Ulrich Ruch and Taskin, 2024, Gosselin and Taskin, 2023, Albrizio et al., 2023, Kalyani and Ozkan, 2025). Appendix E.2 details the corresponding literature for each variable. The questionnaire is primarily designed to capture forward-looking information. Rather than simply extracting historical accounting figures, it identifies management's expectations, pricing directions, and planned capital allocations that may provide leading signals about aggregate economic conditions.

To promote comparability across firms and over time, each substantive variable is constrained to a controlled response format, such as fixed directional categories (e.g., increasing, stable, decreasing), ordinal scales, or bounded numeric values. If a filing does not address a specific topic, the model is instructed to record it as omitted rather than imputing a value. Furthermore, to guarantee data provenance, the model outputs three auxiliary metadata fields for every answer: *evidence* (the verbatim supporting passage), *confidence* (an assessment of textual clarity), and *status* (an indicator of whether the answer is explicitly stated or logically inferred from context). Calibrated on a development sample, these constraints allow me to systematically trace every data point back to its source text and formally distinguish between reported facts and model inferences during robustness checks.

Figure 2: Map of the 38 variables used in the analysis, grouped into eight panels



Notes: Numbers are variable IDs matching the full extraction instrument in Appendix D and Table D.1; document identifiers, narrative fields, and classification variables (IDs 1–5 and 44–50) are extracted but not used in the analysis.

3.4 Prompt Structure

The extraction prompt instructs the LLM to act as a senior macro-financial analyst and to code all 50 questionnaire items for each filing. The model receives the questionnaire together with the corresponding 10-K or 10-Q filing and returns a structured JSON object containing the coded answer, supporting evidence or inference basis, confidence rating, and disclosure status for each variable. To preserve the real-time information structure of the data, the prompt prohibits the use of information released after the filing date and instructs the model to base each answer on information contained in the filing. This restriction is designed to reduce the risk of look-ahead contamination from the model's pretrained knowledge. The full text of the prompt is reported in Figure 3.

Figure 3: Extraction Prompt Template

```
You are a senior financial analyst at the U.S. Federal Reserve conducting structured analysis
of corporate 10-Q and 10-K filings.

TEMPORAL CONSTRAINT: You MUST NOT use any knowledge of events, data, or information that occurred
after {filing_date}. This is critical for accurate historical analysis.

TASK: Read the {company_name} filing and generate a JSON response for all questions from the
questionnaire framework.

FILING LOCATIONS:
- Questionnaire: {questionnaire_abs}
- Filing: {filing_abs}
- Output: {output_abs}

STEPS:
1. Read all questions in the questionnaire.
2. Read the filing.
3. Generate complete JSON for every metric.

JSON STRUCTURE:
{{
  "metric_name": {{
    "question": "[Question from questionnaire]",
    "answer: "[Your answer]",
    "status": "[EXPLICIT/IMPLIED/NOT_MENTIONED/NOT_APPLICABLE]",
    "evidence": "[Supporting text evidence from document, max 300 chars]",
    "confidence": "[HIGH/MEDIUM/LOW]",
    "reasoning": "[Explain logic and financial interpretation behind the answer]",
    "analyst perspective": "[Brief note on analytical viewpoint or significance]",
  }}
}}

FIELD CONSISTENCY:
- Use HIGH, MEDIUM, or LOW exactly.
- IMPLIED answers must include the inference basis in "evidence".

STATUS DEFINITIONS:
- EXPLICIT: Directly stated in the filing.
- IMPLIED: Reasonably inferred from available evidence.
- NOT_MENTIONED: No relevant information; optional fields only.
- NOT_APPLICABLE: Not relevant; optional fields only.
```

```
EVIDENCE RULES:
- EXPLICIT: Exact quote or close paraphrase.
- IMPLIED: Brief explanation of the inference basis.
- Blank only for NOT_MENTIONED or NOT_APPLICABLE.
- Maximum 300 characters.

CRITICAL DATA HANDLING:
- Required fields must never be empty.
- When explicit data is unavailable, infer a reasonable answer using:
  - company industry and business model;
  - historical information available as of {filing_date};
  - economic and peer context as of {filing_date}; and
  - standard business practices.
- Mark inferred answers as "IMPLIED" with confidence "LOW".
- Never use "N/A", "Data unavailable", or equivalent text.

FORMAT STANDARDS:
- Percentages: 1 decimal place, e.g., 5.2%.
- Currency: millions of USD with 1 decimal place.
  Example: $119.575 billion becomes "119575.0".
- Dates: YYYY-MM-DD.
- Dropdowns: Use the exact questionnaire option.
- Multiselects: Return a JSON array.
- Integers: Whole numbers only.
- Apply the questionnaire's specified decimal precision.

EXAMPLE:
{{
  "revenue_growth_guidance_pct": {{
    "Question": "Revenue growth guidance for next period (%)",
    "Answer": "5.2",
    "Reasoning": "Management provided explicit forward guidance supported by product and
                  services growth.",
    "Analyst Perspective": "The guidance indicates continued growth while reflecting
                  macroeconomic uncertainty.",
    "status": "EXPLICIT",
    "evidence": "Management expects approximately 5% revenue growth next quarter.",
    "confidence": "HIGH"
  }}
}}

QUALITY STANDARDS:
- Maintain analytical objectivity.
- Provide clear evidence-to-conclusion links.
- Answer every questionnaire item.
```

The prompt imposes common coding and formatting rules to facilitate comparison across firms and reporting periods. Categorical variables are restricted to prespecified response options, while percentages, monetary values, dates, and other numerical variables follow standardized units and precision rules defined in the questionnaire. These restrictions reduce variation in wording and numerical representation and facilitate construction of the panel dataset.

Each observation is assigned one of four disclosure statuses: EXPLICIT, IMPLIED, NOT_MENTIONED, or NNOT_APPLICABLE. Explicit answers require supporting text from the filing, whereas implied answers must be supported by sufficient contextual information within the filing. NOT_MENTIONED indicates that an item is relevant but not discussed, while NOT_APPLICABLE indicates that it is not relevant to the firm, its business model, or the reporting period. The model also reports a confidence level for each answer. Together, the evidence, status, and confidence fields provide an observation-level audit trail that I use in subsequent validation and robustness analyses.

3.5 From Firm Narratives to Macroeconomic Activity: An Illustrative Example

To illustrate how qualitative firm disclosures capture macro-relevant developments before they are fully reflected in aggregate statistics, we trace Ford Motor Company's SEC filings during the 2021 global semiconductor shortage. The episode shows how firm-level textual signals evolve from emerging supply risks into binding constraints on real output.

An early signal appears in Ford's 2020 Form 10-K, filed February 5, 2021. Beyond quantifying potential production losses, management describes the economic mechanism underlying the emerging bottleneck:

> "Global semiconductor makers allocated more capacity to meet surging demand for consumer electronics during the COVID-19 pandemic [...] At the same time, wafer foundries [...] have not invested enough in recent years to increase capacities to the levels needed to support demand from all of their customers."[4]

The disclosure provides economic context that is difficult to recover from contemporaneous accounting quantities alone. The shortage reflected both a reallocation of semiconductor capacity toward consumer electronics and insufficient upstream capacity as automotive demand recovered. Supplier estimates placed 10–20% of planned first-quarter production at risk, while Ford characterized the situation as sufficiently fluid that the effect was not incorporated into its formal guidance.

Table 1 summarizes the structured economic signals extracted by the LLM. In the February filing, the model classifies the production plan as MAINTAIN, an inferred answer with medium confidence. This interpretation is consistent with Ford's statement that its guidance excluded the semiconductor shortage: the model distinguishes the baseline plan from the potential production shortfall caused by supply constraints. Separately, it identifies operating capacity constraints, continued supply-chain disruption, elevated supply-chain risk, and limited forward visibility.

[4]Ford Motor Company, Form 10-K for fiscal year 2020, Item 1A, *Risk Factors*, p. 16.

Table 1: Evolution of LLM-extracted economic signals: Ford, 2021

Question	**2020 Form 10-K**	**2021Q1 Form 10-Q**
Production/output change plan	Maintain (I, M)	Decrease (E, H)
Operating capacity constraints	Yes (E, H)	Yes (E, H)
Supply-chain status	Still disrupted (E, H)	Worsening (E, H)
Supply-chain risk concern level	4 / 5 (E, H)	5 / 5 (E, H)
Customer demand outlook	Uncertain (E, H)	Strong (E, H)
Inventory relative to demand	Insufficient (I, L)	Insufficient (E, H)
Pricing direction	Increase (E, H)	Increase (E, H)
Forward visibility	Limited (E, H)	Limited (E, M)

Notes: The table reports selected outputs from the common LLM questionnaire applied to Ford's Form 10-K filed February 5, 2021 and Form 10-Q filed April 29, 2021. E and I denote *Explicit* and *Implied* classifications; H, M, and L denote High, Medium, and Low confidence, respectively.

By Ford's 2021Q1 Form 10-Q, filed April 29, 2021, the constraint had materialized and intensified. Semiconductor shortages eliminated approximately 17% of planned first-quarter production, while expected second-quarter losses increased to roughly 50% of plan. More importantly for measuring soft information, management explains why its assessment of the disruption had deteriorated:

> "Our initial outlook was for semiconductor supply chains to remain constrained through the second quarter of 2021 [...] However, the industry faced another setback on March 19, 2021, when Renesas Electronics Corporation, a key supplier [...] experienced a significant fire at its Naka Factory [...] we now believe the automotive semiconductor shortage may not be fully resolved until 2022."[5]

The filing therefore conveys more than a deterioration in contemporaneous production. It reveals a revision in management's assessment of both the severity and expected duration of the disruption.

The narrative also helps distinguish supply constraints from weak demand. Despite declining production, Ford reports strong demand and favorable pricing conditions in North America:

> "[...] first quarter 2021 revenue increased 5%, primarily reflecting robust customer demand for the new product portfolio, higher industrywide net pricing on lower dealer inventories from the semiconductor supply shortage, and favorable mix."[6]

Taken together, these disclosures thus reveal an economic configuration that a single sentiment measure could obscure: demand remained strong and prices increased even as input shortages constrained production and depleted dealer inventories.

Narrative information is also useful for interpreting Ford's own inventory data. Book inventories increased from $10.8 billion to $12.7 billion during the quarter. Viewed in isolation, the increase is ambiguous: inventories may rise because of weak final demand, deliberate stock building, or production frictions. Ford's accompanying disclosure identifies the relevant mechanism:

[5]Ford Motor Company, Form 10-Q for the quarter ended March 31, 2021, Item 2, *Recent Developments*, p. 33.
[6]Ford Motor Company, Form 10-Q for the quarter ended March 31, 2021, Item 2, *North America*, p. 37.

Table 2: Selected LLM outputs from Ford's 2021Q1 Form 10-Q

Panel A: Customer demand outlook	
Question	Customer demand outlook assessment
Classification	STRONG (*Explicit*; High confidence)
Evidence	"[. . .] robust customer demand for the new product portfolio, higher industrywide net pricing on lower dealer inventories [. . .]" (North America); China wholesales increased 85% from a year ago.
Reasoning	Management explicitly describes North American demand as robust, while China wholesales also rebounded sharply, supporting a strong demand classification despite production constraints.
Panel B: Inventory levels relative to demand	
Question	Inventory levels relative to demand
Classification	INSUFFICIENT (*Explicit*; High confidence)
Evidence	"[. . .] higher industrywide net pricing on lower dealer inventories from the semiconductor supply shortage [. . .]" alongside "robust customer demand for the new product portfolio."
Reasoning	Management directly links low dealer inventories to semiconductor shortages in an environment of robust demand, supporting an insufficient inventory classification.
Panel C: Production/output change plan	
Question	Production/output change plan
Classification	DECREASE (*Explicit*; High confidence)
Evidence	Ford expected to lose about 50% of planned 2021Q2 production, compared with a 17% loss in Q1, and approximately 1.1 million wholesale units over the full year.
Reasoning	Management explicitly guides to a substantial production shortfall relative to plan, driven by semiconductor shortages, with 2021Q2 expected to be the trough.
Panel D: Supply-chain normalization status	
Question	Supply-chain normalization status
Classification	WORSENING (*Explicit*; High confidence)
Evidence	Expected production losses rise from 17% of plan in 2021Q1 to about 50% in Q2, with management describing Q2 as the expected trough and noting that the estimated full-year impact had increased substantially.
Reasoning	The sharp deterioration in expected production losses indicates that semiconductor supply conditions were worsening into 2021Q2, despite anticipated improvement later in the year.
Panel E: Pricing direction	
Question	Product/service pricing direction
Classification	INCREASE (*Explicit*; High confidence)
Evidence	"[. . .] higher net pricing (reflecting the strength of our product portfolio and lower incentives in response to reduced dealer stock levels)[. . .]"; net pricing contributed $2.954 billion to Automotive EBIT in 2021Q1.
Reasoning	Management explicitly links higher net pricing to reduced dealer inventories and lower incentives, supporting an increasing pricing classification.

Notes: The table reports selected outputs from the common LLM questionnaire applied to Ford Motor Company's Form 10-Q filed April 29, 2021. *Classification* reports the model answer together with whether the underlying information is stated directly in the filing (*Explicit*) or inferred from related context (*Implied*), and the model's reported confidence. Evidence and reasoning are lightly condensed for presentation; the underlying classifications are unchanged.

> "Finished products [. . .] include vehicles completed but awaiting installation of components affected by the semiconductor supply shortage, after which, they will proceed through an additional quality review process prior to being shipped to our dealers."[7]

This textual evidence therefore changes the interpretation of the observed inventory increase. At least part of Ford's book inventory reflected vehicles awaiting semiconductor components, while dealer inventories remained low relative to demand. This example illustrates how qualitative disclosures complement numerical firm data by identifying the economic mechanism underlying an observed quantity.

Consistent with these disclosures, Table 1 shows that the LLM updates customer demand to *Strong*, inventories relative to demand to *Insufficient*, production plans to *Decrease*, supply-chain conditions to *Worsening*, and keeping prices *Increasing*. Table 2 reports the corresponding detailed outputs. In each case, the model retrieves the relevant evidence from the filing, and the resulting classification is consistent with that evidence. Together, these outputs recover the central economic configuration of the episode: strong demand alongside constrained supply, scarce inventories, and declining expected production.

Subsequent aggregate statistics reveal the same sector-specific disruption. In July 2021, the Federal Reserve reported that manufacturing output increased at a 3.7% annual rate in 2021Q2, while motor vehicle and parts production fell at a 22.5% annual rate amid continuing semiconductor shortages (Board of Governors of the Federal Reserve System, 2021). BEA subsequently reported a decline in private inventory investment led by retail trade inventories, while its industry accounts showed that weakness in retail trade was led by motor vehicle and parts dealers (U.S. Bureau of Economic Analysis, 2021a,b).

This case illustrates the informational mechanism underlying our empirical approach. Accounting data record realized quantities; narrative disclosures help explain why those quantities change, how management interprets current conditions, and how firms expect those conditions to evolve. Crucially, the Ford episode contains opposing signals—strong demand and rising prices alongside constrained production and scarce inventories—that would be difficult to summarize with a single scalar sentiment measure. By converting these multidimensional narratives into structured economic variables, the LLM provides a scalable framework for recovering granular firm-level signals that can be aggregated and incorporated into macroeconomic forecasting models.

3.6 Characteristics of the LLM-Extracted Data

The extraction procedure yields valid structured outputs for all 60,236 filings in the sample, corresponding to 3.01 million item-level responses across the 50-item questionnaire. Of these, approximately 2.29 million responses correspond to the 38 substantive variables used in the subsequent empirical analysis. Appendix F details the subsequent restrictions applied to construct the final estimation sample.

[7] Ford Motor Company, Form 10-Q for the quarter ended March 31, 2021, Item 1, Note 9, *Inventories*, p. 16.

Table 3: Distribution of disclosure status and model confidence

(a) Disclosure status			(b) Model confidence		
Status	*N*	%	Confidence	*N*	%
Explicit	1,632,008	54.2	High	982,470	32.6
Implied	1,370,249	45.5	Medium	877,300	29.1
Not_Available	8,347	0.3	Low	1,150,800	38.2
Other	1,196	<0.1	Other	1,230	<0.1
Total	3,011,800	100.0	Total	3,011,800	100.0

Notes: Item-level responses to the 50-item questionnaire for all 60,236 filings (60,236 × 50 = 3,011,800). Panel (a): the disclosure status the model assigns to each answer. Explicit: stated directly in the filing; Implied: inferred from related context; Not_Available: not mentioned in or not applicable to the filing; Other: an unrecognized status label. Panel (b): the model's self-reported confidence in each answer; Other: missing or unrecognized confidence label. Answers with status Not_Available or Other are treated as missing. Percentages may not sum to 100 because of rounding.

Panel (a) of Table 3 summarizes the disclosure status assigned to each output. Approximately 54.2% of responses are classified as Explicit, indicating that the relevant information is stated directly in the filing, while 45.5% are Implied from related textual context. Only 0.3% are classified as Not_Available. The remaining 1,196 responses (less than 0.1%) carry an unrecognized label and are treated as missing. Panel (b) reports the corresponding confidence distribution: 32.6% of outputs receive High confidence, 29.1% Medium, and 38.2% Low.

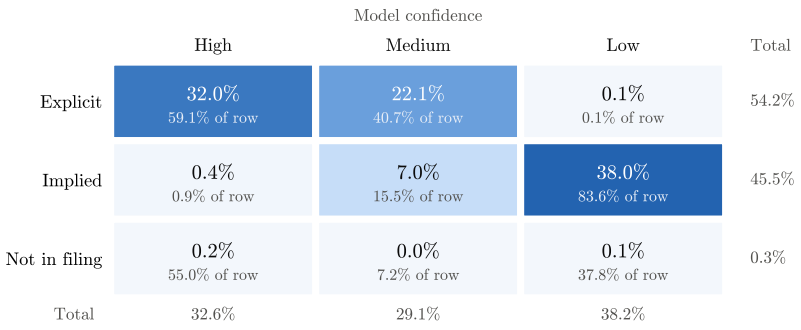


Figure 4: Joint distribution of disclosure status and model confidence

Notes: Large numbers: share of all 3,011,800 item-level responses in each cell. Small numbers: share of the row, that is, the confidence distribution within each disclosure status. Totals are the row and column sums. The 1,233 responses with an unrecognized status or confidence label are omitted.

Disclosure status and model confidence are closely related. Figure 4 reports their joint distribution. Of the Explicit answers, 59.1% are rated High and 40.7% Medium, and only 0.1% Low. Of the Implied answers, by contrast, 83.6% are rated Low, 15.5% Medium and 0.9% High. This pattern is consistent with the prompt design, which requires the model to distinguish direct textual support from contextual inference. These confidence ratings are used primarily to stratify observations in validation and robustness exercises.

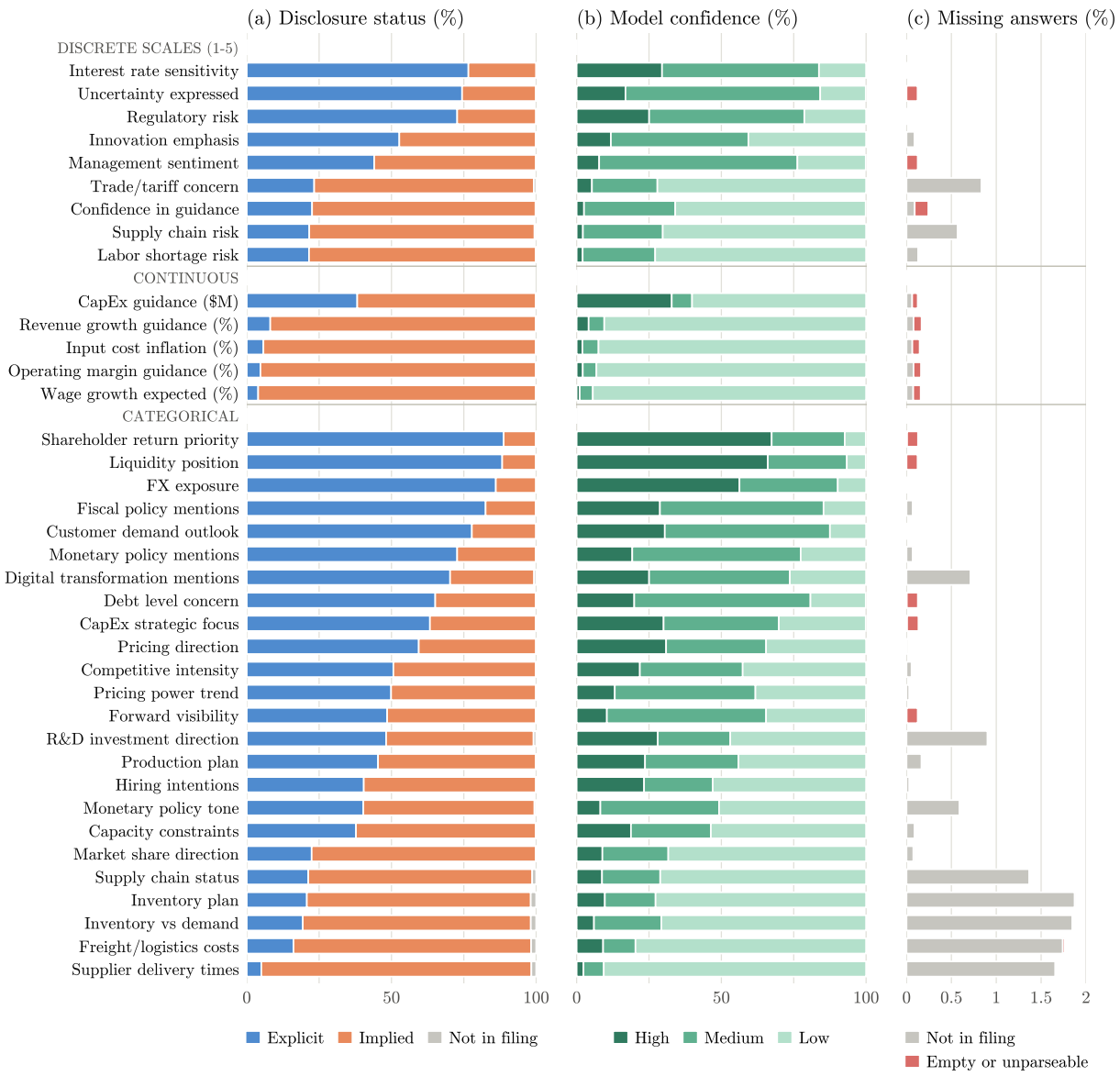


Figure 5: Disclosure status, model confidence, and missing answers by variable

Notes: The 38 variables used in the analysis, 60,236 filings, grouped as in Table 4 and sorted within each group by the share of explicit answers. (a) Share of answers the model classifies as stated in the filing (EXPLICIT), inferred from related context (IMPLIED) or not covered by the filing. (b) The model's self-reported confidence in each answer. (c) Missing answers as a percentage of filings: not covered by the filing, or an empty or unparseable output; note the different scale.

Panels (a) and (b) of Figure 5 reveal substantial heterogeneity across questionnaire items. Variables tied to recurring financial and risk disclosures are frequently stated explicitly. By contrast, forward-looking quantitative variables—including revenue growth, operating margins, wage growth, and input-cost inflation—are predominantly classified as IMPLIED, with explicit disclosures appearing in only 4–8% of filings. Inspection of these responses shows that, when management does not provide an explicit forward-looking estimate, the model often anchors its implied numerical as-

sessment on contemporaneous information reported for the filing period. This pattern reflects an important feature of corporate disclosure: managers typically communicate outlooks qualitatively or through range guidance rather than precise point forecasts for individual operating-cost components (Lansford et al., 2013, Mayew et al., 2026). Accordingly, implied numerical outputs should be interpreted as model-based assessments informed by contemporaneous disclosures rather than as explicit management forecasts.

Panel (c) of Figure 5 decomposes missing responses into genuine non-coverage (NOT_AVAILABLE) and extraction failures (empty or unparseable outputs). Missingness is low for all variables. The largest non-coverage rates, approximately 1.4–1.9%, occur for supply-chain and inventory measures, consistent with differences in industry exposure: these operational topics are less relevant for many financial and software firms. Extraction failures remain below 0.2% for every variable.

3.7 Variable Construction, Timing, and Summary Statistics

To construct aggregate macroeconomic indices from the firm-level textual extractions, I convert categorical and ordinal outputs into numerical measures, retain continuous variables in their original units, and date each observation by the date on which the underlying filing became publicly available.

Categorical variables (24 items). Categorical answers are mapped onto the unit interval $[0, 1]$ with equal steps between the ordered options, so that higher values correspond to more of the economic concept the question measures; for example, debt-level concern runs from NO CONCERN $= 0$ to HIGH CONCERN $= 1$, and the liquidity position from CONCERNING $= 0$ to STRONG $= 1$.

Four questions offer options that do not lie on one axis, and I code each of them on a single scale that keeps only the options on that axis. *Inventory relative to demand*: INSUFFICIENT and EXCESS are opposite directions of one imbalance, so the variable measures the net direction, INSUFFICIENT $= 0$, BALANCED $= 0.5$, EXCESS $= 1$; an even mix of firms with shortages and gluts averages to the balanced value. *Customer demand outlook*: WEAK $= 0$, MODERATE $= 0.5$, STRONG $= 1$; UNCERTAIN (17% of answers) describes confidence rather than strength and is treated as missing. *CapEx strategic focus*: MAINTENANCE $= 0$, EFFICIENCY $= 1/3$, TECHNOLOGY UPGRADE $= 2/3$, CAPACITY EXPANSION $= 1$ measure how expansive the spending is; MIXED (57% of answers) is not a point on this scale and is treated as missing. *Shareholder return priority* measures how much capital is returned: GROWTH/INVESTMENT $= 0$, DIVIDENDS or BUYBACKS $= 0.5$, BOTH $= 1$; the choice between the two channels is not a difference in amount and is not coded. The sector of the firm is a grouping variable, not one of the 38 measures.

Blank, non-applicable and unparseable answers are treated as missing and never replaced by a default value. They account for 0.23% of the answers to the other 36 variables, concentrated in the inventory and supply-chain questions (0.8–1.4%); with the off-axis answers above, 2.2% of all answers are missing.

Discrete scales (9 items). Variables elicited on 1–5 integer scales are first checked for range validity and then linearly rescaled to the unit interval: $\{1 \to 0,\ 2 \to 0.25,\ 3 \to 0.5,\ 4 \to 0.75,\ 5 \to 1\}$. Values outside the admissible 1–5 range are treated as missing.

Continuous variables (5 items). The five continuous variables—revenue growth, operating margin, CapEx, input-cost inflation, and expected wage growth—remain in their original units, expressed as percentages or millions of dollars. These measures include both explicitly disclosed values and LLM-implied assessments when the filing does not provide a direct numerical forecast.

To limit the influence of extreme observations and extraction anomalies without introducing look-ahead bias, each continuous variable is winsorized at the 1st and 99th percentiles computed over a rolling window of the filings in the four calendar quarters before the filing's quarter (all earlier filings at the start of the sample). Thus, the winsorization thresholds applied to a filing depend only on information available before the quarter in which that filing is submitted.

Timing. Each observation enters the dataset on its EDGAR filing date rather than on the end date of the fiscal period to which the filing refers. To construct weekly firm-level series, each firm contributes the answers of its most recent filing, carried forward for at most 52 weeks. The filing is used as it stands: if an answer in the latest filing is missing, the firm does not contribute to that variable, and an older answer is not used in its place. If a firm has not filed for more than one year, it is treated as missing. No observations are backfilled or interpolated, ensuring that information from later filings cannot affect earlier dates.

Summary statistics. Table 4 reports summary statistics for all 38 quantitative variables over the full sample of 60,236 filings from 1,159 firms between January 1996 and September 2026. The statistics are computed after invalid values are removed but before the continuous variables are winsorized.

Panel A summarizes the nine 1–5 discrete scales. The extracted management-sentiment measure averages 3.53, while confidence in guidance averages 3.23. Trade-policy concern and supply-chain risk average 1.75 and 2.15, respectively. Panel B reports the continuous variables. Mean extracted revenue growth is 4.87%, while expected wage growth averages 3.51%. The raw distributions contain substantial tails: revenue growth reaches the prompt bounds of −50% and 100%, operating margins reach −100% and 100%, and CapEx reaches $230 billion. These extremes motivate the real-time winsorization described above. Panel C reports the categorical variables after numerical coding. The liquidity-position measure averages 0.86, the inventory measure averages 0.53, close to balanced with a slight tilt toward excess inventory, and operating capacity constraints are identified in 28% of observations. The final column reports the share of non-missing answers classified as EXPLICIT. Explicit disclosure rates vary substantially across variables, ranging from approximately 4–5% for expected wage growth and operating-margin measures to 89% for shareholder-return priority.

Table 4: Summary statistics of the extracted answers (before winsorization)

ID	Variable	N	Mean	Std	Min	Max	% explicit
Panel A: Scales (1–5)							
24	Supply chain risk	60,229	2.15	0.79	1	5	21
25	Regulatory risk	60,231	3.38	0.80	1	5	73
26	Labor shortage risk	60,226	2.09	0.67	1	5	21
27	Interest rate sensitivity	60,229	3.10	0.92	1	5	77
32	Trade/tariff concern	60,228	1.75	0.85	1	5	23
34	Innovation emphasis	60,228	3.09	1.02	1	5	53
40	Management sentiment	60,159	3.53	0.68	1	5	44
41	Confidence in guidance	60,140	3.22	0.61	1	5	23
42	Uncertainty expressed	60,163	3.44	0.71	1	5	75
Panel B: Continuous variables							
6	Revenue growth guidance (%)	60,141	4.87	8.50	−50	100	8
7	Operating margin guidance (%)	60,146	21.93	18.55	−100	100	5
8	CapEx guidance ($M)	60,164	1,296	4,362	0	230,000	38
9	Input cost inflation (%)	60,156	3.24	4.08	−50	50	6
10	Wage growth expected (%)	60,154	3.51	1.49	−30	30	4
Panel C: Categorical variables (coded 0–1)							
11	Inventory plan	59,389	0.58	0.29	0	1	21
12	Pricing direction	60,217	0.67	0.37	0	1	59
13	Production plan	60,162	0.79	0.33	0	1	45
14	Hiring intentions	60,222	0.57	0.39	0	1	41
15	Capacity constraints	60,226	0.28	0.45	0	1	37
16	Customer demand outlook	49,957	0.62	0.33	0	1	76
17	Market share direction	60,222	0.64	0.30	0	1	22
18	Competitive intensity	60,205	0.77	0.26	0	1	50
19	Pricing power trend	60,225	0.39	0.36	0	1	50
20	Supply chain status	59,733	0.87	0.25	0	1	21
21	Supplier delivery times	59,691	0.47	0.17	0	1	5
22	Freight/logistics costs	59,661	0.61	0.26	0	1	16
23	Inventory vs demand	59,478	0.53	0.17	0	1	20
28	FX exposure	60,228	0.55	0.41	0	1	86
29	Monetary policy mentions	60,227	0.46	0.31	0	1	73
30	Monetary policy tone	59,957	0.42	0.27	0	1	40
31	Fiscal policy mentions	60,227	0.58	0.30	0	1	82
33	R&D investment direction	59,940	0.72	0.29	0	1	48
35	Digital transformation mentions	60,218	0.37	0.35	0	1	70
36	CapEx strategic focus	25,980	0.69	0.37	0	1	68
37	Liquidity position	60,163	0.86	0.17	0	1	88
38	Debt level concern	60,163	0.22	0.21	0	1	65
39	Shareholder return priority	60,121	0.65	0.38	0	1	89
43	Forward visibility	60,160	0.49	0.19	0	1	49

Notes: The sample contains 60,236 filings from 1,159 firms between January 1996 and September 2026, restricted to filings submitted while the firm was a member of the S&P 500. N denotes the number of non-missing answers; blank, unrecognized, and not-applicable responses are treated as missing. Panel A reports integer scores from 1 to 5, with values outside $[1, 5]$ removed. Panel B reports continuous variables in their original units before winsorization. Panel C reports categorical variables coded on $[0, 1]$ as described in Section 3.7. The final column reports the share of non-missing answers classified as EXPLICIT; remaining valid answers are classified as IMPLIED. ID denotes the question number in the extraction questionnaire reported in the Appendix D.

3.7.1 Do the Answers Measure Firm Conditions?

To assess whether the LLM-extracted signals capture economically meaningful firm conditions, I test whether they predict subsequent realized firm outcomes. Adapting the firm-level predictive framework of Gennaioli et al. (2016), I merge the filing data with Compustat through CRSP and estimate

$$\text{outcome}_{i,t\to t+4} = b \cdot \text{answer}_{i,t} + c \cdot \text{sales growth}_{i,t-4\to t} + \alpha_i + \tau_t + \varepsilon_{i,t}. \tag{1}$$

Firm fixed effects α_i absorb time-invariant differences across firms, while calendar-quarter fixed effects τ_t absorb aggregate conditions common to firms in a given quarter. Consequently, identification comes from within-firm variation in the extracted signal, relative to other firms observed in the same quarter and conditional on recent sales growth. The coefficient b therefore measures whether changes in a firm's extracted condition are systematically associated with its subsequent realized outcomes.

I map the 38 extracted signals to 12 accounting outcomes: growth in sales, capital expenditures, R&D, employment, staff expenses, and inventories; changes in operating and gross margins; and changes in leverage, cash holdings, shareholder payouts, and the interest cost of debt. Six policy-related measures lack a direct accounting counterpart and are evaluated against subsequent sales growth. Of the 60,236 filings in the full sample, 52,709 can be linked to Compustat, and 50,861 enter the main validation regressions after requiring the relevant outcomes and controls.

Table 5 reports the primary validation results. Twenty-eight of the 38 extracted signals significantly predict their preassigned outcome at the 5% level. Demand- and production-related measures are especially informative about subsequent sales growth. A one-standard-deviation increase in the extracted revenue-growth measure is associated with a 0.45-standard-deviation increase in realized sales growth over the following four quarters ($t = 29.64$). Customer demand, production plans, and market-share direction yield t-statistics between 15.85 and 16.70, while greater expressed uncertainty predicts weaker sales growth ($t = -10.40$). The relationships extend to other outcomes: hiring intentions predict employment growth ($t = 13.09$), pricing power predicts operating-margin changes ($t = 11.28$), R&D investment direction predicts R&D growth ($t = 8.97$), and inventory plans predict inventory growth ($t = 9.63$).

The weakest results are concentrated among continuous forward-looking measures for which explicit numerical disclosure is uncommon. Operating-margin, CapEx, and input-cost measures do not reliably predict their corresponding outcomes, while expected wage growth does not predict staff-expense growth. Among policy-related measures, monetary-policy tone ($t = 3.93$), regulatory risk ($t = -3.82$), and exchange-rate exposure ($t = -2.06$) are significantly associated with subsequent sales growth.

The final columns of Table 5 examine whether the predictive relationships are driven solely by explicitly stated information. Among the 28 measures that predict their preassigned outcomes in the full sample, 22 remain significant using IMPLIED observations alone. The implied revenue-growth measure, production plans, and market-share direction remain particularly predictive, with

Table 5: Firm-level predictive validation of extracted measures

Extracted measure	Full-sample t	Std. effect	Implied t	Explicit–implied diff. t	Explicit share
Sales growth					
Revenue growth	29.64	0.450	29.69	−6.71	0.09
Market share	16.70	0.167	17.06	−5.15	0.22
Production plan	16.24	0.220	15.85	0.34	0.46
Demand outlook	15.85	0.250	8.52	1.90	0.76
Sentiment	13.57	0.201	12.82	2.27	0.45
Confidence in guidance	11.22	0.125	9.88	2.67	0.23
Forward visibility	10.27	0.125	9.15	0.99	0.48
Uncertainty	−10.40	−0.133	−8.42	−1.07	0.74
CapEx growth					
Capacity constraints	8.26	0.073	5.26	1.55	0.39
CapEx focus	2.64	0.054	1.92	1.58	0.69
CapEx	−1.24	−0.022	−1.45	1.05	0.40
R&D growth					
R&D direction	8.97	0.112	0.38	1.26	0.92
Innovation	3.62	0.057	3.82	−1.22	0.79
Digital/AI mentions	−1.03	−0.015	0.71	−1.58	0.68
Employment growth					
Hiring intentions	13.09	0.152	9.89	0.36	0.40
Labor shortage	4.17	0.048	3.97	−0.34	0.22
Staff-expense growth					
Expected wage growth	−0.04	−0.001	0.24	−0.38	0.09
Operating margin					
Pricing power	11.28	0.136	7.59	−0.98	0.50
Operating margin	−2.12	−0.034	−2.29	1.44	0.05
Competition	−8.79	−0.079	−7.28	2.63	0.50
Gross margin					
Pricing direction	6.06	0.082	1.39	2.04	0.59
Freight costs	1.73	0.017	2.06	−1.51	0.16
Input-cost inflation	0.58	0.009	0.93	−1.22	0.06
Supply-chain status	−0.31	−0.004	−0.30	−0.14	0.22
Supply-chain risk	−0.96	−0.014	−1.61	1.22	0.22
Inventory growth					
Inventory plan	9.63	0.111	8.91	−1.30	0.25
Supplier delivery	−5.91	−0.055	−5.93	0.22	0.06
Inventory vs. demand	−14.46	−0.125	−12.81	0.42	0.23
Leverage					
Debt concern	−13.31	−0.121	−7.56	−2.45	0.65
Cash / assets					
Liquidity	−13.07	−0.112	−4.32	−3.64	0.88
Payout / assets					
Shareholder returns	−2.03	−0.018	−1.63	1.12	0.89
Interest cost / debt					
Rate sensitivity	12.43	0.185	8.35	0.69	0.76
Policy environment (no direct firm outcome): sales growth					
Monetary-policy tone	3.93	0.031	2.77	−0.63	0.40
Monetary-policy mentions	0.37	0.003	0.68	−0.46	0.72
Fiscal mentions	0.01	0.000	1.22	−1.27	0.82
Trade-policy concern	−0.08	−0.001	−1.30	1.08	0.21
FX exposure	−2.06	−0.036	−1.89	0.62	0.86
Regulatory risk	−3.82	−0.036	−1.49	−2.48	0.73

Notes: The table reports estimates from equation (1). Regressions include firm and calendar-quarter fixed effects and trailing four-quarter sales growth. Standard errors are two-way clustered by firm and calendar quarter. Continuous answers and outcomes are winsorized at the 1st and 99th percentiles. The estimation sample contains 50,861 filings for the main outcomes. *Std. effect* equals the estimated coefficient multiplied by the standard deviation of the extracted measure and divided by the standard deviation of the outcome. *Implied* reports the t-statistic on the extracted measure when IMPLIED observations form the baseline group. *Explicit–implied difference* reports the t-statistic on the interaction between the extracted measure and an indicator for EXPLICIT status; it therefore tests whether the predictive slope differs between explicit and implied observations. The final column reports the share of non-missing observations classified as EXPLICIT. Outcomes are measured from fiscal quarter t through $t + 4$.

t-statistics of 29.69, 15.85, and 17.06, respectively. The predictive slopes for explicit and implied observations differ for some variables, but not systematically.

Table 6: Firm-level validation across alternative outcomes

Answer	Sales	Capex	R&D	Empl.	Staff	Op. m.	Gr. m.	Inv.	Lev.	Cash	Payout	Int.
Revenue growth guidance	**29.6**	**17.6**	**13.8**	**9.6**	**6.1**	**9.7**	**7.3**	**10.1**	**−4.7**	**−5.7**	**8.9**	**2.4**
Market share	**16.7**	**11.1**	**8.4**	**9.2**	**4.1**	**4.4**	**3.4**	**8.5**	−1.1	**−5.9**	**5.3**	0.0
Production plan	**16.2**	**13.2**	**7.7**	**9.3**	**5.5**	**4.6**	1.7	**10.8**	−0.1	**−6.8**	**5.7**	**−2.0**
Demand outlook	**15.8**	**15.9**	**7.8**	**8.7**	**2.8**	**7.9**	**4.8**	**9.3**	−1.7	**−5.5**	**9.5**	**−3.2**
Sentiment	**13.6**	**17.9**	**9.5**	**11.0**	**4.6**	**6.2**	**4.1**	**9.3**	0.5	**−8.7**	**11.2**	**−8.2**
Confidence in guidance	**11.2**	**14.0**	**6.8**	**9.1**	1.5	**5.1**	**3.2**	**8.4**	1.5	**−5.4**	**7.4**	**−6.8**
Forward visibility	**10.3**	**12.3**	**6.9**	**8.9**	**2.0**	**4.0**	1.2	**6.4**	1.7	**−4.9**	**6.8**	**−6.0**
Uncertainty	**−10.4**	**−12.6**	**−6.6**	**−10.1**	−1.9	**−3.8**	−1.2	**−7.2**	**−2.4**	**4.9**	**−7.4**	**7.0**
Capacity constraints	**5.9**	**8.3**	**3.6**	1.6	−0.4	**3.4**	**3.6**	**5.4**	−0.3	**−4.2**	**2.7**	−1.3
Capex focus	**3.7**	**2.6**	**4.6**	**3.3**	**2.2**	**−2.7**	**−2.1**	**4.0**	**2.0**	−1.5	1.0	0.7
Capex guidance ($)	**−3.5**	−1.2	−0.1	**−3.8**	1.4	**−3.2**	−1.9	−1.2	**4.2**	−1.7	−1.4	**2.5**
R&D direction	**5.2**	**7.3**	**9.0**	**2.1**	1.4	**−3.2**	−0.7	1.9	**2.9**	**−2.5**	0.7	−0.8
Innovation	**4.4**	**4.9**	**3.6**	1.4	0.8	−1.5	−0.7	**3.1**	**2.2**	−1.1	1.4	−1.1
Digital/AI mentions	0.3	0.0	−1.0	**−2.5**	−0.7	−1.2	−1.7	1.5	1.3	−0.2	**−2.3**	0.5
Hiring intentions	**13.5**	**14.0**	**10.6**	**13.1**	**6.0**	0.7	0.6	**10.5**	1.8	**−7.0**	**4.7**	**−3.5**
Labor shortage	**8.3**	**8.3**	**5.9**	**4.2**	0.3	1.2	1.7	**5.6**	1.0	**−3.0**	1.6	−1.5
Wage growth expected	**7.1**	**8.2**	**7.7**	**4.6**	−0.0	−0.6	0.3	**4.6**	1.4	**−4.6**	**3.0**	**−2.7**
Pricing power	**11.1**	**12.4**	**5.9**	**5.6**	1.4	**11.3**	**9.8**	**6.2**	**−2.5**	**−4.2**	**8.9**	**−4.2**
Margin guidance	**4.1**	**6.7**	**4.2**	**5.6**	0.9	**−2.1**	**−2.7**	**5.3**	1.6	**−2.9**	1.0	**−3.1**
Competition	**−7.7**	**−9.4**	**−3.7**	**−6.8**	−1.2	**−8.8**	**−8.1**	**−4.4**	0.6	**3.8**	**−7.6**	**3.0**
Pricing direction	**9.6**	**9.0**	**4.0**	**3.3**	1.8	**8.4**	**6.1**	**5.1**	**−2.4**	−1.5	**6.1**	**−3.1**
Freight costs	**4.5**	**3.3**	−0.3	−0.4	1.3	**3.1**	1.7	1.8	0.0	−0.0	−0.6	−1.5
Input-cost inflation	**5.5**	**5.0**	**2.7**	0.2	−0.2	**2.2**	0.6	**2.8**	−1.8	−1.5	0.8	−0.0
Supply-chain status	−1.5	**−2.5**	−0.3	1.7	0.3	0.7	−0.3	−1.6	0.8	0.7	**−2.1**	0.2
Supply-chain risk	1.1	1.3	−0.2	−1.6	−0.5	−1.1	−1.0	1.1	−0.1	−1.6	1.1	0.6
Inventory plan	**12.2**	**11.6**	**7.4**	**6.7**	**2.1**	**4.6**	2.0	**9.6**	**2.2**	**−7.1**	**4.6**	**−2.1**
Supplier delivery	**−5.3**	**−5.2**	**−3.3**	−1.3	−1.0	−1.3	0.3	**−5.9**	0.3	**2.8**	−0.8	0.9
Inventory vs demand	**−10.3**	**−11.1**	**−5.2**	**−5.0**	**−2.3**	**−6.2**	**−5.0**	**−14.5**	**2.7**	**5.8**	**−6.7**	**6.5**
Debt concern	−0.3	**−7.6**	**−4.6**	**−10.0**	−1.6	**2.7**	**2.5**	**−5.8**	**−13.3**	**4.0**	**−10.9**	**17.8**
Liquidity	**2.6**	**8.6**	**5.9**	**8.3**	1.1	−1.3	−0.9	**6.8**	**7.9**	**−13.1**	**11.3**	**−12.2**
Shareholder returns	**−7.2**	−1.5	−1.8	0.1	−0.7	**−6.5**	**−6.4**	0.5	**9.8**	−0.8	**−2.0**	**−7.5**
Rate sensitivity	0.2	**−4.6**	**−3.7**	**−6.9**	0.1	1.3	0.6	**−7.8**	**−11.4**	1.9	**−9.1**	**12.4**
Monetary-policy tone	**3.9**	**5.1**	**3.7**	**3.6**	1.4	−1.6	**−2.2**	**3.7**	**4.1**	**−2.2**	**5.9**	**−8.3**
Monetary-policy mentions	0.4	0.5	−0.6	−1.7	−1.1	−1.3	−1.1	−1.7	**−3.3**	0.8	**−2.7**	**3.6**
Fiscal mentions	0.0	0.5	0.5	−0.1	0.3	**2.5**	**3.3**	−1.0	1.1	−1.8	1.0	**−2.1**
Trade-policy concern	−0.1	−0.7	1.2	−0.3	−1.0	−0.4	−0.7	0.1	−0.6	0.8	−1.1	−0.5
FX exposure	**−2.1**	−0.1	0.2	**−4.4**	−1.3	**−2.3**	−1.7	−1.7	−1.5	**3.7**	−1.2	**2.4**
Regulatory risk	**−3.8**	**−3.5**	−1.6	**−5.0**	−1.6	0.6	1.6	**−2.6**	−1.2	1.3	−0.8	**3.0**

Notes: Each cell reports the t-statistic on the extracted measure from a separate regression of the indicated subsequent firm outcome, using the specification in equation (1). Bold entries satisfy $|t| \geq 1.96$. Underlined entries denote the preassigned primary validation outcome used in Table 5. Outcomes are sales, CapEx, R&D, employment, and staff-expense growth; changes in operating and gross margins; inventory growth; leverage; cash/assets; payout/assets; and interest cost/debt. Staff expense is available for approximately 8,100 filings and R&D expenditure for approximately 14,400 filings.

These results indicate that implied answers contain economically meaningful information about firms' underlying conditions rather than merely adding classification noise. I therefore retain both

explicit and implied observations in the baseline index construction. As a broader diagnostic, Table 6 reports the full 38×12 outcome matrix; because it involves many comparisons, I treat it as supporting rather than primary validation evidence.

4 Aggregate Filing Measures: Construction and Validation

4.1 Weekly Aggregation and Time-Series Properties

Because corporate filings arrive asynchronously according to firms' fiscal calendars, I first convert the firm-level extractions into weekly time series. Figure 6 illustrates the resulting irregularity in the information flow for 2025. Filing arrivals are strongly clustered within each quarter: over 1996–2026, a typical week contains roughly 13 filings, while reporting deadlines generate spikes of more than 300 filings in the busiest weeks. Nevertheless, coverage is nearly continuous, with only about 1% of sample weeks containing no filings.

I construct weekly aggregates using the timing conventions in Section 3.7. On each Friday, every S&P 500 constituent contributes its most recent non-missing answer filed during the preceding 52 weeks, and each variable is aggregated using an equal-weighted mean across contributing firms. Answers enter only on their EDGAR filing dates and are never backfilled. Hence, each weekly observation uses only information that was publicly available by that date.

Figure 7 plots the resulting 38 aggregate series together with NBER-dated recessions. A visible feature of several raw aggregates is quarterly filing-form seasonality. Annual reports tend to contain more extensive discussion of some topics than quarterly reports, generating recurring differences between 10-K and 10-Q observations. For example, the average 10-K response exceeds the average 10-Q response by 0.21 for digital-transformation mentions, 0.14 for fiscal-policy mentions, and 0.11 for competitive intensity on their $[0, 1]$ scales, and by 0.13 for innovation emphasis on its 1–5 scale. As annual reports enter the panel and are subsequently replaced by quarterly filings, these differences generate a pronounced saw-tooth pattern in some series. Beyond this mechanical variation, the aggregate series exhibit several distinct economic patterns.

Cyclical indicators. Measures of demand, production, hiring, and managerial expectations are strongly procyclical. For example, the customer demand outlook index averages 0.68 during the 2005–2007 expansion but collapses to 0.35 by December 2001, 0.10 by May 2009, and 0.19 in June 2020. Production plans, hiring intentions, management sentiment, and confidence in guidance mirror this trajectory closely. Interestingly, forward visibility reaches 0.12 in June 2020, compared with a trough of approximately 0.32 during the Global Financial Crisis. This suggests the pandemic induced an unprecedented forecasting friction, even though aggregate demand fell less severely than during the Global Financial Crisis. Expressed uncertainty moves in the opposite direction, rising from a 2005–2007 average of 3.27 to 4.17 in March 2009 and 4.70 in June 2020. These patterns indicate that the textual measures capture substantial variation in firms' assessments of current conditions and the outlook over the business cycle.

Secular trends. Several variables instead display pronounced low-frequency movements. Digital-

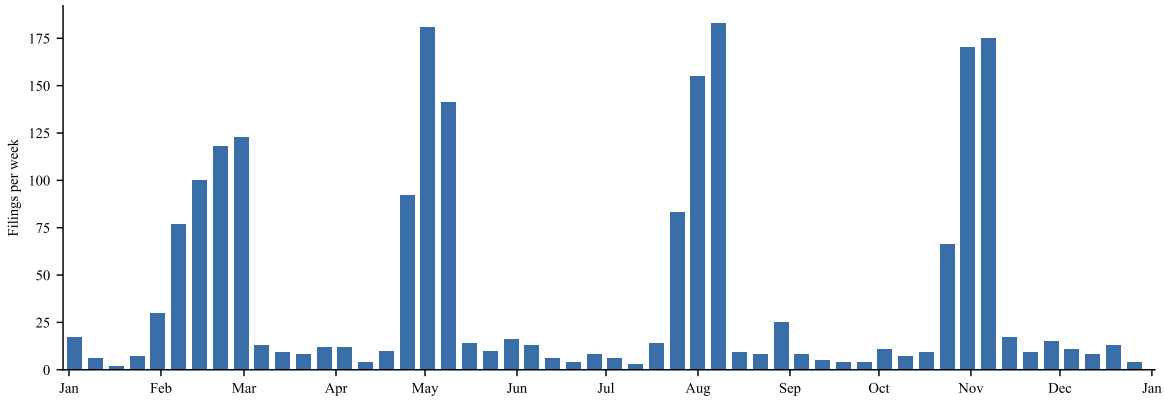


Figure 6: Weekly filing arrival rate, 2025

Notes: Each bar is the number of filings (10-Q and 10-K) by S&P 500 members in a week ending on Friday, by EDGAR filing date, 2025. The spikes follow the filing deadlines, about 40–45 days after the end of each fiscal quarter for 10-Qs and 60–90 days after the fiscal year for 10-Ks.

transformation mentions rise from approximately 0.17 in 1997 to 0.50 by 2025, while innovation emphasis increases from about 2.8 to 3.3 on its 1–5 scale. Nominal CapEx rises from roughly $580 million to $2.3 billion per firm over the sample. Because this measure is expressed in nominal dollars, the increase reflects a combination of firm growth, inflation, changing sample composition, and investment behavior.

Event-driven measures. Certain variables respond primarily around particular macroeconomic or geopolitical shocks. Concern regarding trade policy averages 1.5 in the mid-2000s, steps up to 2.1 during the 2018–2019 trade war, and sharply peaks at 3.2 following tariff escalations in early 2026. Supply-side measures move sharply during the post-pandemic disruption: supply-chain risk, labor-shortage risk, capacity constraints, and freight costs all rise substantially and reach local or sample peaks around 2021–2022. Capacity constraints, for example, rise to approximately 0.70 in early 2022. Input-cost inflation reaches 7.3% in August 2022, while expected wage growth peaks at 4.9% later that year. Policy-related measures also display episodic movements, with fiscal-policy mentions increasing around major fiscal-policy changes and monetary-policy mentions rising during the 2022–2023 tightening cycle.

Slow-moving measures. Finally, several aggregates exhibit relatively limited time-series variation. Liquidity, foreign-exchange exposure, interest-rate sensitivity, market-share direction, and regulatory risk move considerably less than the more cyclical or event-driven variables. These series appear to capture relatively persistent features of firm conditions rather than high-frequency business cycle variation. The stability of these series confirms that the LLM extraction and aggregation methodology does not mechanically generate spurious cyclicality where underlying firm disclosures remain static.

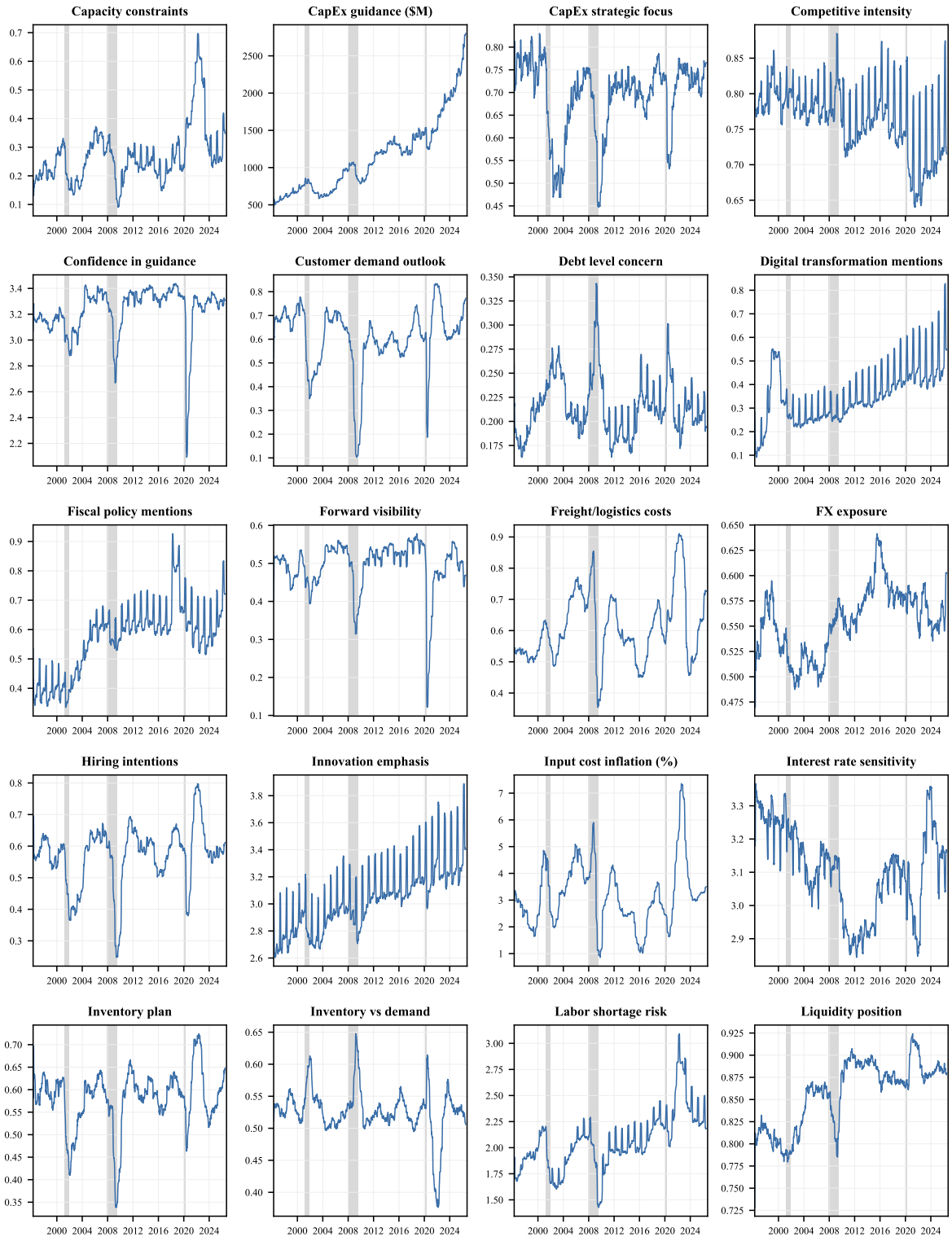


Figure 7: Weekly aggregates of extracted filing variables (part 1)

Notes: Weekly aggregates of the first 20 quantitative answers (in alphabetical order), 1996–2026. Each Friday, the equal-weighted mean over S&P 500 members of each firm's latest answer filed in the previous 52 weeks. Scales as in Table 4. Gray bands mark NBER recessions.

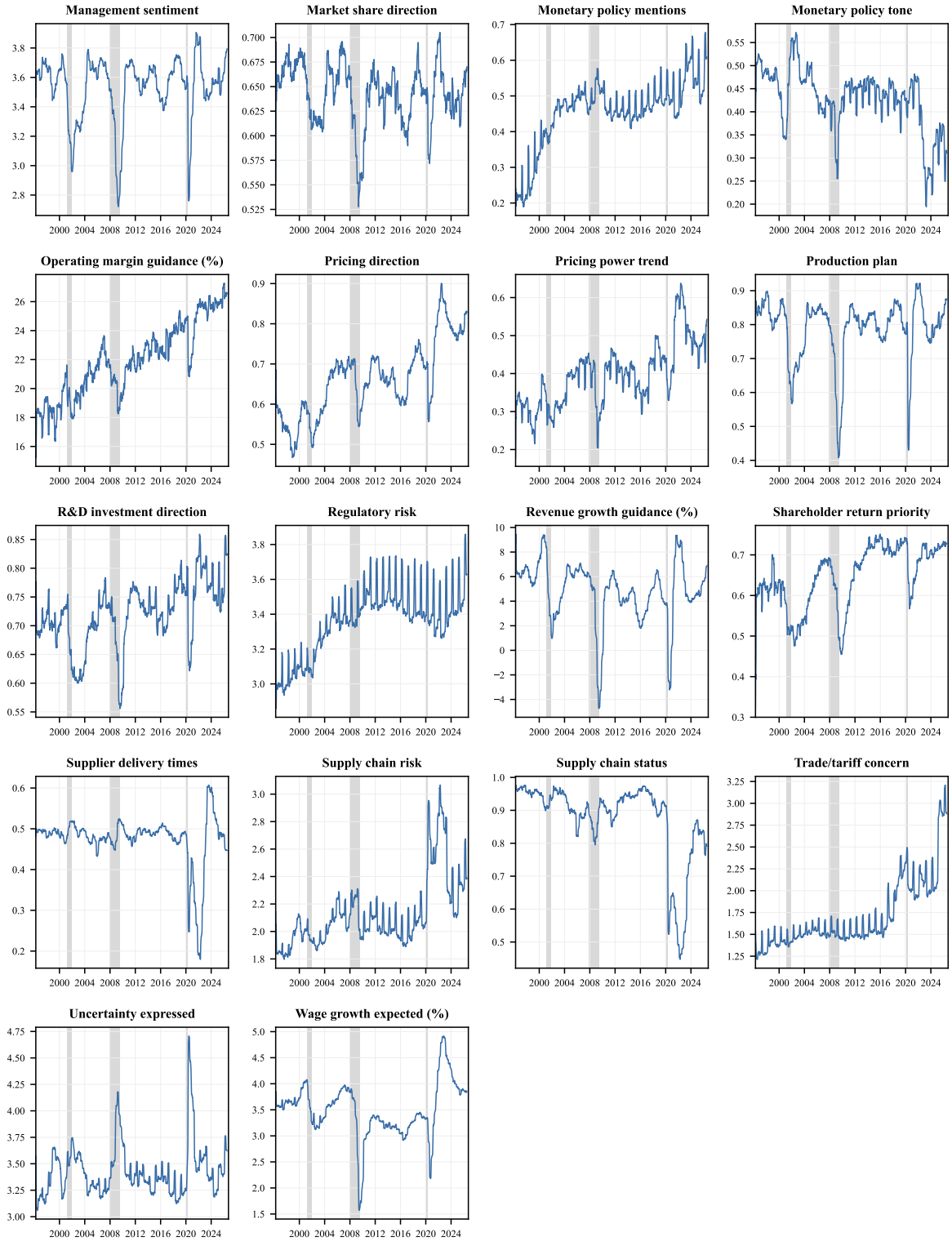


Figure 7: Weekly aggregates of extracted filing variables (part 2)

Notes: Weekly aggregates of the remaining 18 quantitative answers (in alphabetical order), 1996–2026. Each Friday, the equal-weighted mean over S&P 500 members of each firm's latest answer filed in the previous 52 weeks. Scales as in Table 4. Gray bands mark NBER recessions.

4.2 External Validation Against Macroeconomic Benchmarks

I next assess whether the filing-based aggregates capture the economic conditions they are intended to measure. I compare 12 representative filing aggregates with independently constructed macroeconomic benchmarks selected based on conceptual correspondence. For comparability, I conduct the analysis at the quarterly frequency: weekly filing aggregates are averaged within calendar quarters, and monthly benchmark series are similarly averaged within quarters, while benchmarks observed directly at the quarterly frequency are used as reported. Table 7 reports correlations in levels and in four-quarter changes, and Figure 8 plots the corresponding series.

The real-activity measures closely track conventional indicators over the business cycle. Customer demand outlook comoves strongly with expected sales growth from the Survey of Business Uncertainty (SBU), with a correlation of 0.90. Production plans and hiring intentions also move closely with industrial production and payroll growth, with correlations of 0.78 and 0.73, respectively. The relationships are not confined to persistent differences in levels: the corresponding four-quarter-change correlations range from 0.73 to 0.90. More importantly, the visual correspondence is economically sensible. Production plans weaken around the 2001 recession, contract sharply during the Global Financial Crisis, and fall again in 2020, mirroring contractions in realized industrial activity. Hiring intentions display a similar pattern relative to payroll growth. These comparisons suggest that managers' descriptions of planned production and hiring contain information about the same underlying cyclical conditions that are reflected in official activity data.

Inventory plans provide another particularly close comparison. The filing-based measure and business inventory growth move together through both expansions and contractions, including the sharp inventory adjustment during the Global Financial Crisis and the large swings surrounding the pandemic. Their correlation is 0.86. This relationship is economically intuitive: firms adjust planned inventories in response to expected sales, production needs, and existing stocks, so common movements in the two series indicate that the qualitative disclosures capture firms' inventory-adjustment cycle rather than only general sentiment.

The extracted price and cost measures also correspond closely to aggregate inflation dynamics. Input-cost inflation rises and falls with intermediate-goods PPI inflation, including the commodity-price movements surrounding the Global Financial Crisis and the pronounced increase in input costs in 2021–2022. Expected wage growth similarly tracks wage inflation measured by the Employment Cost Index, with a correlation of 0.79. The timing is not exact, however: turning points in the filing measure and the ECI do not always occur in the same quarter. Pricing power shows an even stronger relationship with core PCE inflation ($\rho = 0.79$). In particular, firms' reported ability to raise prices increases markedly during the 2021–2022 inflation episode and subsequently moderates as inflation declines. This pattern is consistent with pricing power reflecting firms' ability to pass higher costs through to customers rather than simply recording contemporaneous price changes.

The filing measures also capture economically distinct supply constraints. Supply-chain status moves inversely with the Global Supply Chain Pressure Index ($\rho = -0.72$): firms report sharply deteriorating supply conditions during the pandemic disruption and substantial normalization there-

Table 7: Correlation of filing aggregates with macroeconomic benchmarks

Dimension	Filing aggregate	External benchmark	Exp. sign	Correlation	
				Levels	4-qtr changes
Demand	Customer demand outlook	SBU expected sales growth	+	0.90	0.90
Production	Production plan	Industrial production (y/y)	+	0.78	0.81
Labor demand	Hiring intentions	Payrolls (y/y)	+	0.73	0.73
Inventories	Inventory plan	Business inventories (y/y)	+	0.86	0.88
Input prices	Input cost inflation (%)	PPI, intermediate (y/y)	+	0.62	0.53
Wages	Wage growth expected (%)	ECI wages and salaries (y/y)	+	0.79	0.76
Pricing	Pricing power trend	Core PCE inflation (y/y)	+	0.79	0.74
Supply chains	Supply chain status	GSCPI	−	−0.72	−0.47
Labor shortages	Labor shortage risk	Vacancies / unemployed	+	0.89	0.77
Financial conditions	Debt level concern	NFCI	+	0.53	0.55
Trade policy	Trade/tariff concern	Trade policy uncert.	+	0.76	0.55
Uncertainty	Uncertainty expressed	SBU sales uncertainty	+	0.80	0.90

Notes: Pearson correlations of quarterly averages, 1996Q2–2026Q2 or from the first quarter both series exist (GSCPI from 1998Q1, vacancies from 2000Q4, ECI from 2002Q1, SBU from 2016Q4). Each filing aggregate is paired with one benchmark that measures the same concept; the expected sign is the sign implied by the two definitions (the NFCI rises when financial conditions tighten, the GSCPI when supply chains are under pressure). Levels: correlation of the series; 4-quarter changes: correlation of $x_q - x_{q-4}$, which removes common trends. Benchmarks: FRED; Federal Reserve Bank of Atlanta Survey of Business Uncertainty (SBU); Federal Reserve Bank of New York (GSCPI, Benigno et al., 2022); Caldara et al. (2020) (trade policy uncertainty).

after. Labor-shortage risk, meanwhile, closely follows the ratio of vacancies to unemployed workers ($\rho = 0.89$). Both measures rise as labor markets tighten during the late 2010s and especially during the 2021–2022 recovery, before easing as labor market tightness declines. The extraction distinguishes product- and input-market bottlenecks from labor-market constraints rather than producing a single generic measure of "difficulty" or negative sentiment.

The remaining comparisons illustrate that the extracted information extends beyond conventional measures of real activity. Debt-level concern rises during periods of financial stress and is positively correlated with the Chicago Fed National Financial Conditions Index ($\rho = 0.53$). The relationship is most visible around episodes such as the Global Financial Crisis, when both financial conditions and corporate balance-sheet concerns deteriorated sharply. The correlation is more modest during normal periods, which is consistent with debt concerns becoming particularly informative when financing conditions bind rather than moving one-for-one with aggregate financial conditions at all times.

Trade and tariff concern exhibits a different type of variation. The measure is relatively subdued for much of the early sample but rises strongly alongside the Trade Policy Uncertainty index during periods of elevated trade-policy risk, yielding a correlation of 0.76. Unlike production or employment, this variable is therefore better interpreted as an event-sensitive measure: it becomes salient when trade policy moves to the center of firms' operating environment. Similarly, uncertainty expressed in corporate filings closely tracks SBU sales uncertainty over their common sample ($\rho = 0.79$), with both series rising sharply around the pandemic. This comparison is particularly informative because uncertainty is inherently qualitative and is not directly observable from firms' accounting quantities.

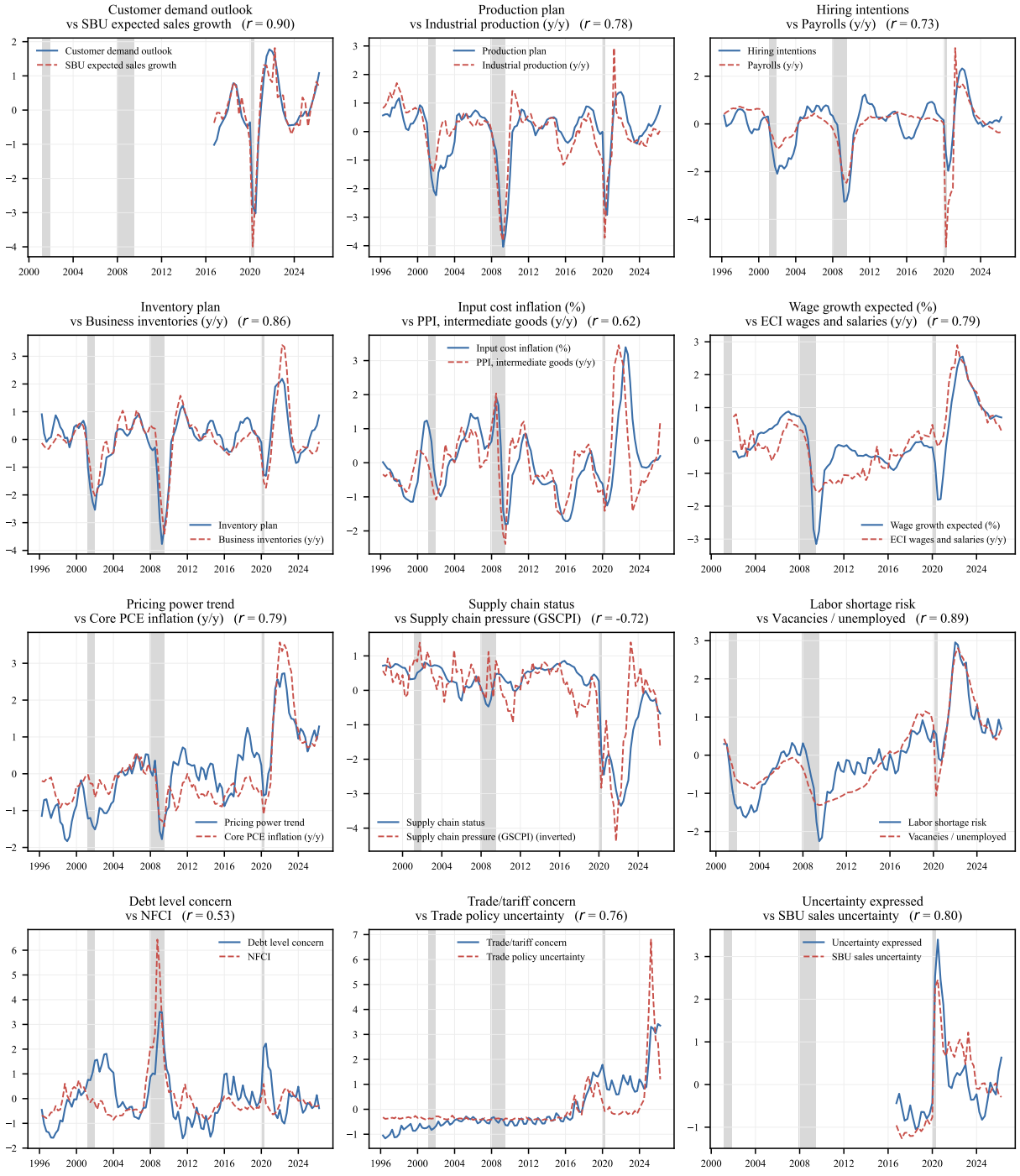


Figure 8: External validation of filing-based aggregate measures

Notes: Standardized quarterly averages (z-scores) of twelve filing aggregates (solid) and their benchmarks (dashed), 1996Q2–2026Q2 or from the first quarter both series exist (GSCPI from 1998, vacancies from 2000, ECI from 2002, SBU from 2016Q4). The GSCPI, which has an expected negative sign, is plotted inverted. r is the correlation in levels from Table 7. Gray bands mark NBER recessions.

Taken together, the benchmark comparisons provide external validation for the economic content of the filing-based measures. The correlations are strongest when the external series closely matches the underlying questionnaire concept. Moreover, many relationships remain strong in four-quarter changes, indicating that the correlations are not driven simply by common long-run trends. The filing-based measures are not intended to reproduce existing indicators. They are constructed from a common firm-level information source, are available across a much broader set of economic dimensions, and can be aggregated jointly into the macroeconomic indices developed below.

5 GDP Nowcasting and Forecasting with Filing-Based Measures

Corporate filings improve short-horizon forecasts of first-release real GDP growth in the pre-pandemic evaluation period. With variables selected on 1996–2012 and held fixed, a bridge equation combining separate leading and coincident factors reduces mean squared forecast errors relative to an autoregressive benchmark by 17 percent for the current quarter and 18 percent one quarter ahead, averaged over weeks 5–13 of 2013–2019. These predictors also improve quantile forecasts, with larger relative gains under a loss function that emphasizes downside risk. These gains largely disappear after 2020. This section details the variable selection, factor construction, and recursive forecasting procedure used to obtain these results.

5.1 Real-Time Design and Information Sets

Target. The forecasting target is year-over-year real GDP growth as first published in the BEA advance estimate. For each quarter, I use the ALFRED vintage of real GDP in which that observation first appears, evaluating models against the exact information set available in real time rather than subsequent statistical revisions.

Filing panels. The filing sample follows Section 3.7: 60,236 Form 10-K and 10-Q filings from 1,159 firms between 1996 and 2026, dated by their EDGAR filing dates and included only while the firm is a member of the S&P 500. For the weekly panels, firms are assigned to one of $S = 11$ sectors. For variable j in sector s and week t, the sector aggregate is

$$X_{s,j,t} = \frac{1}{|C_{s,j,t}|} \sum_{i \in C_{s,j,t}} x_{i,j,t}, \tag{2}$$

where i indexes firms. The value $x_{i,j,t}$ is the filing-based measure j for firm i available in week t. The set $C_{s,j,t}$ contains firms in sector s contributing an available observation for that measure, and $|C_{s,j,t}|$ is their number. Thus, $X_{s,j,t}$ is the equal-weighted sector mean. The stock and flow panels differ in which observations contribute.

I use three related data structures. For variable selection, I construct an economy-wide quarterly panel: each firm's latest observation is assigned to the calendar quarter in which the filing becomes public, carried forward for at most four quarters, and averaged across firms. For the weekly stock

panel, used to construct continuous indices, each firm contributes its latest filing provided that it is no more than 52 weeks old. Finally, the weekly flow panel uses only firms that file in the current week. The flow panel therefore measures newly arriving filing information rather than the stock of previously disclosed information. Many filings become public after the advance GDP release for the fiscal quarter they cover (Figure 6). Their potential forecasting value therefore lies in managerial assessments and operating plans that reveal ongoing conditions and intended adjustments in production, hiring, and investment.

Levels and changes. The level and the change of a filing-based measure contain different macroeconomic information. A level describes the state of a condition—for example, whether demand is strong or firms are actively hiring—whereas a change records whether the aggregate assessment has strengthened or weakened. Changes may be informative about turning points, although they also reflect changes in the composition of contributing firms. I therefore allow each of the 38 filing-based measures to enter either in levels or as a quarter-on-quarter change. Once one form of a variable is selected, the other is removed from the candidate set. In the weekly stock panel, changes are measured over 13 weeks; in the flow panel, they are measured relative to the firm's previous filing. Section 7.3 considers alternative transformation rules.

5.2 Selecting Leading and Coincident Filing-Based Measures

All variable selection uses the 1996Q2–2012Q4 training sample.

Lead–lag patterns. I begin by characterizing the timing of each filing-based measure relative to GDP:

$$Y_q = \alpha + \beta_{j,k} X_{j,q-k} + u_q, \tag{3}$$

where Y_q is first-release year-over-year real GDP growth. The regressor $X_{j,q-k}$ is the quarterly filing aggregate, entered in levels or changes in the corresponding regression. Positive k means that the filing-based measure precedes GDP growth by k quarters; negative k means that it follows. The intercept is α, and $\beta_{j,k}$ is the slope coefficient for measure j at horizon k, and u_q is the regression error term. Newey–West standard errors use bandwidth four. Appendix Tables G.1 and G.2 report the full results.

Levels of demand, production, sentiment, and inventories tend to be coincident or lagging, whereas changes in many of the same measures precede GDP growth. Changes in revenue guidance, inventories relative to demand, wage expectations, and hiring plans are informative one or more quarters ahead. Supply-chain risk is an exception: its level precedes weaker activity. These training-sample associations motivate separate leading and coincident candidate pools; the forecasting exercise below evaluates whether the selected signals retain predictive content after 2012.

Predictive selection. I next select a set of leading and coincident variables using pseudo-out-of-sample performance within the training period. The 38 variables in levels and changes generate

76 candidate series. In the predictive regressions, $\widetilde{X}_{v,q}$ denotes a standardized candidate series, which may be the level or change of a filing-based measure; v indexes candidate series in these regressions, while j still indexs the underlying economic measures. In forward selection, standardization moments are recomputed from each recursive estimation sample and applied to its forecast observation.

A series enters the leading candidate pool if its largest absolute correlation with Y_{q+h} over $h = 1, \ldots, 8$ is at least 0.30 and is at least 95 percent of its absolute contemporaneous correlation. A coincident candidate instead has its largest absolute correlation at $h = 0$, with a value of at least 0.30. These rules yield 41 leading and 16 coincident candidates; Appendix G.1 reports the full cross-correlograms.

For the leading candidates, I first evaluate marginal predictive content using

$$Y_{q+1} = \alpha + \rho Y_{q-1} + \beta_v \widetilde{X}_{v,q} + u_{q+1}, \tag{4}$$

where Y_{q-1} is the latest GDP observation that would have been available at the end of quarter q, ρ is its autoregressive coefficient, and β_v measures the candidate's conditional predictive association with next-quarter growth. Predictive performance is measured relative to the autoregressive benchmark using the Campbell–Thompson out-of-sample R^2 on an expanding window.

I then apply forward selection. Starting from an empty set S_L, each step adds the variable that minimizes pseudo-out-of-sample MSFE:

$$Y_{q+1} = \alpha + \rho Y_{q-1} + \sum_{m \in S_L} \gamma_m \widetilde{X}_{m,q} + \beta_v \widetilde{X}_{v,q} + u_{q+1}. \tag{5}$$

Here S_L is the set of already selected leading series, m indexes its members, γ_m is the coefficient on selected series m, and $v \notin S_L$ indexes the additional candidate being evaluated. The first 24 quarters form the initial estimation window. Selection stops when no remaining candidate improves the out-of-sample R^2 or when five variables have been selected.

The resulting leading set is

$$S_L = \left\{ \begin{array}{l} \Delta\,\text{Revenue growth guidance}, \quad \text{Supply-chain risk}, \\ \Delta\,\text{Inventory relative to demand}, \quad \Delta\,\text{Expected wage growth}, \quad \Delta\,\text{Hiring intentions} \end{array} \right\}.$$

Table 8 reports the selection paths. The pseudo-out-of-sample R^2 rises from 0.603 with the first selected variable to 0.754 with all five. Four of the five variables enter as changes. Increases in revenue guidance and hiring intentions are consistent with stronger expected sales and labor demand. A shift toward excess inventories predicts weaker growth, consistent with unwanted accumulation when demand falls short of plans. Supply-chain risk enters negatively, as expected if input shortages constrain production, while stronger wage expectations accompany stronger activity.

I construct the coincident set analogously, replacing the one-quarter-ahead target with contemporaneous GDP growth and excluding alternative transformations of variables already selected for

S_L. The procedure yields

$$S_C = \{\text{Uncertainty}, \quad \text{R\&D investment direction}, \quad \text{Production plan}, \quad \Delta\,\text{CapEx guidance}\}\,.$$

The associated pseudo-out-of-sample R^2 rises from 0.10 with the first selected variable to 0.55 with all four. Production and investment measures summarize operating conditions, while expressed uncertainty enters with a negative sign. Table 8 reports both selection paths.

Table 8: Variable selection on the training sample (1996Q2–2012Q4)

Step	Variable added	Pseudo-out-of-sample R^2 vs. AR
Leading set S_L: 41 CCF leading candidates, target Y_{q+1}		
1	Revenue growth guidance (Δ)	0.603
2	Supply-chain risk	0.706
3	Inventory vs demand (Δ)	0.719
4	Wage growth expected (Δ)	0.741
5	Hiring intentions (Δ)	0.754
Coincident set S_C: 16 CCF coincident candidates, target Y_q		
1	Uncertainty	0.102
2	R&D direction	0.173
3	Production plan	0.500
4	Capex guidance \$ (Δ)	0.549

Notes: Forward stepwise selection inside 1996Q2–2012Q4. Each step adds the candidate that maximizes the pseudo-out-of-sample R^2 of the marginal regression (5) relative to the AR benchmark, computed on an expanding window within the training sample (first 24 quarters as the initial window); selection stops when no candidate improves the R^2 or five variables are selected. Δ marks first-differenced variables.

5.3 Weekly Leading and Coincident Indices

Factor model. I summarize each selected set with a one-factor dynamic factor model. After standardizing the component series, the model is

$$Z_{r,t} = \lambda_r f_t + e_{r,t}, \qquad e_{r,t} \sim N(0, \sigma_r^2), \tag{6}$$

$$f_t = \phi f_{t-1} + \eta_t, \qquad \eta_t \sim N(0, \sigma_\eta^2). \tag{7}$$

Here r indexes selected sector–variable combinations, and $Z_{r,t}$ is the corresponding standardized series in its selected level or change form. The latent factor f_t captures common variation, λ_r is its loading on component r, and $e_{r,t}$ is the component-specific disturbance with variance σ_r^2. The coefficient ϕ governs factor persistence, and η_t is the factor innovation with variance σ_η^2. The model is estimated by the EM algorithm, with the Kalman filter accommodating the unbalanced weekly panel (Bańbura et al., 2011, Bańbura and Modugno, 2014). Estimation is initialized using principal components. I leave the idiosyncratic components serially uncorrelated because the sparse observation pattern makes richer idiosyncratic dynamics difficult to identify.

I estimate the model separately on the weekly stock panels for the leading and coincident sets.

Before estimation, each component is oriented so that higher values indicate stronger economic conditions: supply-chain risk and inventory relative to demand are reversed in the leading set, and uncertainty is reversed in the coincident set. Each factor is then rescaled to have mean 50 and standard deviation 10. The leading index, I_t^L, combines five variables across 11 sectors, for 55 component series; the coincident index, I_t^C, combines four variables across 11 sectors, for 44 component series.

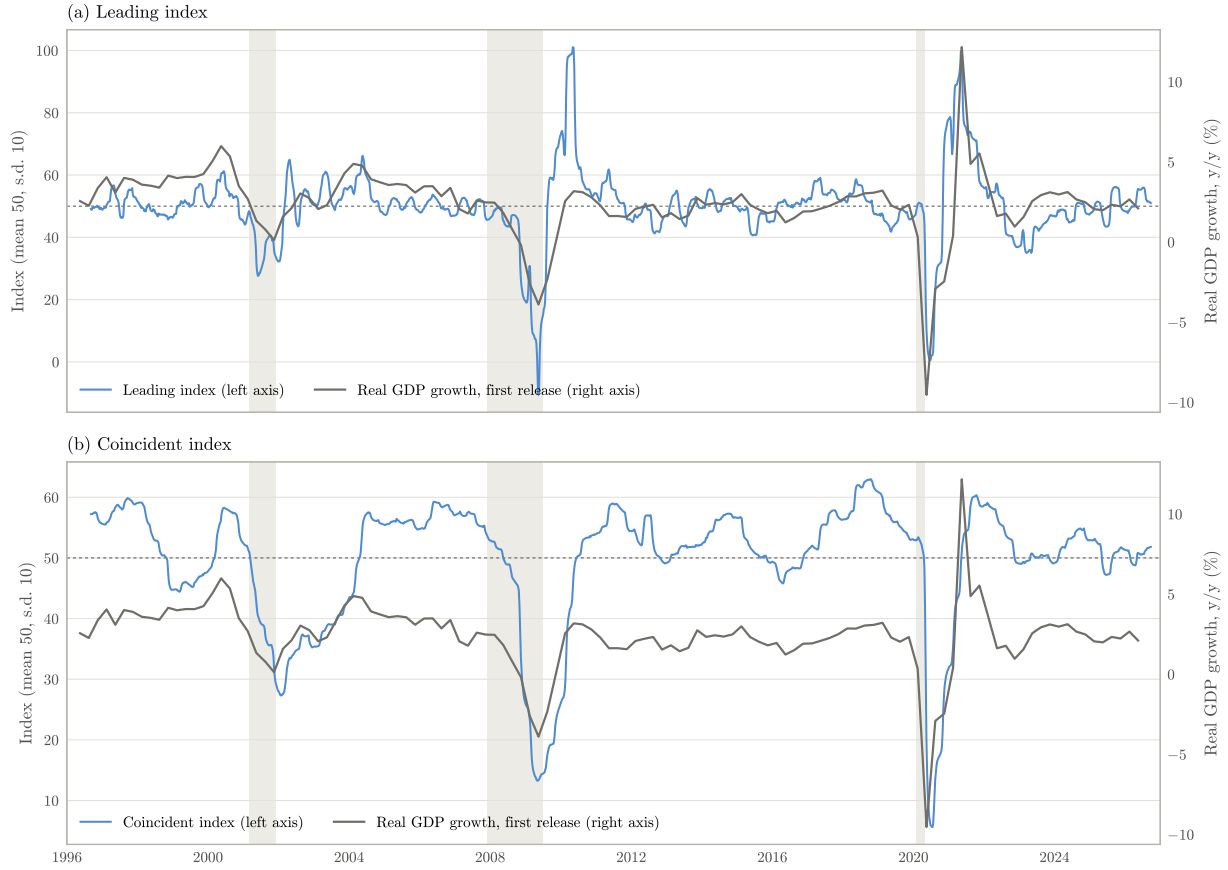


Figure 9: Weekly filing-based leading and coincident indices and real GDP growth as first published.
Notes: Variables are selected using 1996–2012 data. For this descriptive figure, factor-model parameters and standardization moments are estimated over 1996–2019 and the filter is extended through 2026. Shaded regions denote NBER recessions; the dashed line denotes the index mean of 50.

Figure 9 plots I_t^L and I_t^C alongside real GDP growth, where standardization and model parameters are estimated over 1996–2019 and the filter is extended through 2026. Both decline during NBER recessions, but their timing differs. The leading index weakens ahead of the 2001 downturn, recovers earlier following the 2008–2009 and 2020 contractions, and turns down before first-release GDP growth during the 2015–2016 industrial slowdown. The coincident index reaches its trough closer to GDP. Because these plotted series use parameters estimated through 2019 on variables selected through 2012, they characterize cyclical co-movement rather than real-time predictive performance; the forecasting exercises below instead re-estimate the factor models recursively using

only information available at each forecast date.

Table 9 reports correlations between the quarterly indices at quarter t and first-release year-over-year real GDP growth at quarter $t+h$ over 1996:Q3–2019:Q4. The correlation profiles confirm that the leading and coincident indices capture distinct timing horizons. The filing coincident index peaks contemporaneously at $h = 0$ ($r = 0.611$) and decays rapidly, falling to 0.135 by $h = 2$. In contrast, the filing leading index peaks one quarter ahead at $h = 1$ ($r = 0.624$, compared with 0.580 at $h = 0$) and maintains positive correlation through $h = 4$. At $h = 0$, the filing coincident index tracks first-release GDP growth slightly more closely than both the ADS business conditions index (0.590) and the NFCI (-0.528). At $h \geq 1$, the ADS and NFCI exhibit stronger unconditional correlations and slower decay, reflecting the persistence of hard macroeconomic indicators and financial conditions relative to firms' near-term operating assessments.

Table 9: Correlation of quarterly indices with real GDP growth (YoY, first release) at horizon h quarters

Index	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
Filing leading	0.580	**0.624**	0.400	0.231	0.122
Filing coincident	**0.611**	0.433	0.135	-0.046	-0.162
ADS	0.590	**0.797**	0.673	0.534	0.374
NFCI	-0.528	**-0.649**	-0.494	-0.387	-0.260

Notes: Bold = peak absolute correlation for each index. Positive h: index at t correlated with GDP at $t+h$. The filing indices are the full-sample (1996–2026) versions of Section 5.3 (parameters estimated on 1996–2019); NFCI is negatively correlated with GDP (tighter financial conditions signal weaker growth). Sample: 1996 Q3–2019 Q4 (94 quarters).

5.4 Real-Time Point Forecasts

I evaluate whether filing-based measures improve predictions of first-release real GDP growth over 2013–2019. Measures and transformations are selected using data through 2012 and held fixed during evaluation. At each weekly forecast date, a bridge regression combines two filing-based indices with the latest published GDP observation. Factor models are updated quarterly and forecasting regressions weekly, using the information available at the time. I compare the resulting predictions with autoregressive and random-walk benchmarks.

Forecast timing and information. On each of the first 13 Fridays of quarter q, I produce a nowcast of current-quarter growth ($h = 0$) and forecasts one and two quarters ahead ($h = 1, 2$). The nowcast is therefore updated throughout the quarter, alongside the forecasts for subsequent quarters. Let $w \in \{1, \ldots, 13\}$ denote the forecast week and Y_{q+h} first-release year-over-year real GDP growth. Forecasts before February 2013 are excluded because variable selection uses the advance estimate of 2012Q4 GDP.

The available GDP history changes when the BEA publishes a new advance estimate. Before the

release for quarter $q-1$, the latest observation is Y_{q-2}; thereafter, it is Y_{q-1}. I write this observation as Y_{q-L}, where $L = 2$ before the release and $L = 1$ afterward, following the actual release calendar.

Filing-based inputs. The coincident input uses the weekly flow panel to capture newly disclosed operating conditions. The leading input uses the stock panel to summarize firms' previously disclosed outlooks. I estimate a separate dynamic factor model for each set of selected measures, disaggregated across 11 sectors. At the start of quarter q, both models and their standardization moments are re-estimated using an expanding sample from 1996 through the preceding quarter and then held fixed within q.

Let $\widehat{f}^{a,(q)}_{t|t}$ denote the standardized, filtered estimate for calendar week t, where $a \in \{C, L\}$ identifies the coincident or leading factor and (q) identifies the quarterly parameter vintage. Historical estimates are recomputed under this vintage, and the coincident factor is filtered forward as new filings arrive. Writing $t_{q,\ell}$ for the calendar week of the ℓth forecast Friday in quarter q, I construct the bridge inputs as

$$\begin{aligned} F^C(q,w) &= \frac{1}{w}\sum_{\ell=1}^{w} \widehat{f}^{C,(q)}_{t_{q,\ell}|t_{q,\ell}}, \\ F^L_{q-1} &= \frac{1}{13}\sum_{\ell=1}^{13} \widehat{f}^{L,(q)}_{t_{q-1,\ell}|t_{q-1,\ell}}. \end{aligned} \tag{8}$$

The coincident average $F^C(q,w)$ updates weekly, whereas the previous-quarter leading average F^L_{q-1} remains fixed throughout quarter q. The latter's quarterly parameter vintage is suppressed on the left-hand side.

Bridge equation and benchmarks. At each forecast date, I estimate

$$Y_{q+h} = \alpha + \rho Y_{q-L} + \beta_C F^C(q,w) + \beta_L F^L_{q-1} + u_{q+h}, \qquad h \in \{0,1,2\}. \tag{9}$$

The intercept is α, ρ is the autoregressive coefficient, β_C and β_L are the predictive coefficients on the two filing-based indices, and u_{q+h} is the error term. Regressions are estimated separately for each horizon and forecast week, with these coefficient subscripts suppressed, using only GDP outcomes published by the forecast date. Thus, the leading input is fixed within the quarter, but its estimated coefficient can change.

The autoregressive benchmark omits the filing-based indices,

$$Y_{q+h} = \alpha + \rho Y_{q-L} + u_{q+h},$$

with coefficients estimated separately from the bridge. The random-walk benchmark sets

$$\widehat{Y}^{\mathrm{RW}}_{q+h|q,w} = Y_{q-L}.$$

Both benchmarks use the same GDP information available at the forecast date.

Forecast accuracy. For each horizon and forecast week, relative accuracy is measured by

$$R_{h,w} = \frac{\mathrm{MSFE}_{\mathrm{filing}}(h,w)}{\mathrm{MSFE}_{\mathrm{benchmark}}(h,w)} = \frac{\sum_{q \in \mathcal{Q}_{h,w}} \left(Y_{q+h} - \widehat{Y}^{\mathrm{filing}}_{q+h|q,w}\right)^2}{\sum_{q \in \mathcal{Q}_{h,w}} \left(Y_{q+h} - \widehat{Y}^{\mathrm{benchmark}}_{q+h|q,w}\right)^2}. \tag{10}$$

The set $\mathcal{Q}_{h,w}$ contains the forecast-origin quarters evaluated for both models, and $\widehat{Y}^{m}_{q+h|q,w}$ is model m's forecast made in week w of quarter q. Each MSFE divides the corresponding sum of squared errors by $|\mathcal{Q}_{h,w}|$, which cancels in the ratio. Values below one favor the filing-based model.

Reported significance uses two-sided Diebold–Mariano (DM) tests with Newey–West bandwidth four (Diebold and Mariano, 1995). For comparisons with the nested AR benchmark, I also report the one-sided Clark–West adjusted test (Clark and West, 2007), identifying it separately. The main evaluation ends before the pandemic; Section 5.4 examines subsequent performance.

Forecast performance. Table 10 and Figure 10 show that the two-factor bridge has lower MSFE than the AR benchmark in every forecast week for the current quarter and one quarter ahead. Averaging the week-specific ratios over weeks 5–13 gives reductions of 17 and 18 percent, respectively. At the 10 percent level, the DM test favors the bridge in seven of nine weeks at each horizon; the one-sided Clark–West test favors it in all nine. Gains two quarters ahead are weaker: the average reduction is 2 percent, with significance in two weeks under DM and three under Clark–West. Relative to the random walk, improvements are larger and significant at all three horizons throughout weeks 5–13.

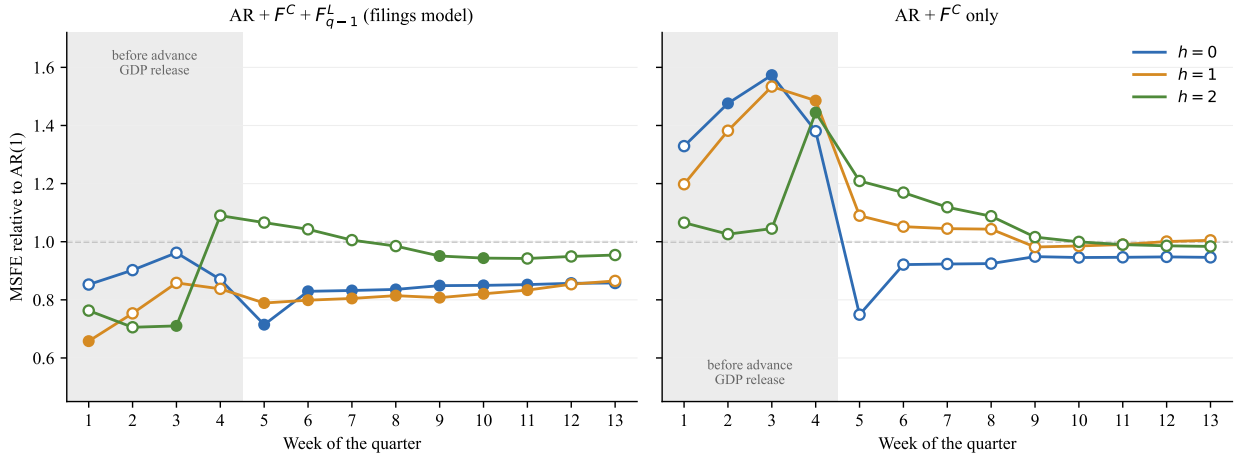


Figure 10: GDP Forecast Accuracy over the Quarter

Notes: MSFE ratios relative to the AR benchmark by forecast week, 2013–2019. The left panel reports the two-factor bridge equation (9); the right panel uses the coincident flow factor only. Filled markers denote rejection of equal forecast accuracy by a two-sided Diebold–Mariano test at 10 percent; a rejection favors the filing model only when its MSFE ratio is below one. Shading marks weeks 1–4, an approximate window before the previous quarter's advance GDP release. The benchmark lag follows the actual release date in each quarter.

Table 10: Real-Time GDP Forecast Accuracy Relative to Benchmark Models

	vs. AR(1)			vs. Random Walk		
Week	$h=0$	$h=1$	$h=2$	$h=0$	$h=1$	$h=2$
$w=1$	0.85	0.66*	0.76	0.65	0.41***	0.39***
$w=2$	0.90	0.75	0.71	0.69	0.47***	0.36***
$w=3$	0.96	0.86	0.71**	0.73	0.53**	0.36***
$w=4$	0.87	0.84	1.09	0.76	0.60*	0.67
$w=5$	0.71***	0.79***	1.07	0.57**	0.55***	0.62**
$w=6$	0.83**	0.80***	1.04	0.76***	0.59***	0.63**
$w=7$	0.83**	0.80***	1.01	0.76***	0.60***	0.61**
$w=8$	0.84*	0.81***	0.98	0.77***	0.61***	0.60**
$w=9$	0.85*	0.81**	0.95*	0.76**	0.61***	0.58**
$w=10$	0.85*	0.82*	0.94*	0.76**	0.62**	0.58**
$w=11$	0.85*	0.83*	0.94	0.77**	0.63**	0.58**
$w=12$	0.86	0.85	0.95	0.77**	0.65**	0.58**
$w=13$	0.86	0.87	0.95	0.77**	0.66**	0.59**

Notes: Each cell reports $\text{MSFE}_{\text{model}}/\text{MSFE}_{\text{benchmark}}$; values below one indicate the model outperforms. Model: equation (9), with the flow-panel coincident factor $F^C(q,w)$ and the leading index averaged over quarter $q-1$. All three specifications use the latest published GDP figure Y_{q-L}, with L set by the actual BEA release calendar. Target: real GDP growth (year over year) as first published. Forecasts made in 2013Q1–2019Q4, excluding dates before February 2013. The evaluation includes 27, 26, and 25 forecast-origin quarters at $h=0,1,2$ in weeks 1–4, and 28, 27, and 26 in weeks 5–13; variables fixed from the 1996–2012 selection; DFMs re-estimated every quarter and bridge equations every week on an expanding window. Diebold–Mariano significance (Newey–West HAC, bandwidth 4): $^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.10$.

In weeks 1–4, the average reductions relative to the AR benchmark are 10, 22, and 18 percent at $h=0,1,2$, but DM significance is limited to one week at each of the latter two horizons. The nowcast improvement is most pronounced in week 5, when its MSFE ratio falls to 0.71, before settling between 0.83 and 0.86 in weeks 6–13. These weekly patterns span changes in both filing availability and the published GDP information set. The previous quarter's advance estimate was available by week 4 in 12 of 28 quarters and by week 5 in 26 quarters; the regressions follow the actual release dates.

Comparing the two panels of Figure 10 shows the benefit of adding the lagged leading index. The coincident flow factor alone performs worse than the AR benchmark in weeks 1–4 at both $h=0$ and $h=1$. In weeks 5–13, its average MSFE ratio is 0.92 for the nowcast and 1.02 one quarter ahead, with no significant improvement under the DM test. Adding the leading index brings the early-quarter ratios below one and reduces one-quarter-ahead MSFE by approximately 20 percent relative to the coincident-only specification in weeks 5–13. A single factor combining all nine measures in a flow panel also performs worse than the AR benchmark at both horizons.

The concentration of statistically significant gains from week 5 onward reflects the institutional timing of the corporate reporting cycle and its interaction with the two-factor specification. As illustrated by the 2025 filing distribution in Figure 6, Form 10-Q and 10-K disclosures are sparse during the first four weeks of a calendar quarter and cluster tightly starting in week 5 ahead of SEC statutory deadlines. Because the coincident input $F^C_{q,w}$ is extracted from the weekly flow panel,

early-quarter estimates rely on a thin cross-section of predominantly off-cycle filers and exhibit substantial sampling noise—explaining why the flow-only specification in the right panel of Figure 10 underperforms the AR benchmark during weeks 1–4. Once the main reporting wave commences in week 5, rapid cross-sectional accumulation across sectors averages out idiosyncratic firm noise, transforming $F^C_{q,w}$ into a precise real-time gauge of current-quarter operating conditions just as the BEA's advance release of Y_{q-1} updates the autoregressive anchor. Meanwhile, the predetermined leading stock factor F^L_{q-1} stabilizes forecasts during the sparse early weeks and provides a broad baseline of previously disclosed corporate outlooks against which incoming filing flows are evaluated, enabling the combined model to significantly improve both current-quarter nowcasts ($h = 0$) and one-quarter-ahead forecasts ($h = 1$).

Comparison with conventional indicators. Table 11 compares the filing-based bridge with specifications using ADS, CFNAI, NFCI, their combination, or the median Survey of Professional Forecasters (SPF) forecast. Each specification retains the latest published GDP observation. ADS, CFNAI, and NFCI use latest-vintage data with assumed publication lags of 7, 25, and 5 days, respectively. The comparisons therefore approximate release timing but incorporate subsequent revisions. The SPF forecast enters as a predictor in the bridge.

In 2013–2019, all reported average MSFE ratios for the conventional-indicator bridges exceed one. In weeks 5–13, the ratios range from 1.09 to 1.24 for the nowcast and from 1.09 to 2.02 one quarter ahead, compared with 0.83 and 0.82 for the two-index filing-based bridge. Adding ADS, CFNAI, and NFCI jointly to the filing-based indices raises these ratios to 1.24 and 1.14. Thus, the filing-based indices provide useful predictive information in this sample, while the additional conventional predictors do not improve accuracy within the bridge specification.

Table 11: Bridge equations with established indicators, forecasts made 2013Q1–2019Q4

	$h = 0$		$h = 1$	
Regressors	weeks 1–4	weeks 5–13	weeks 1–4	weeks 5–13
AR + ADS	1.60	1.11	1.79	1.09
AR + CFNAI	1.22	1.24	1.74	1.69
AR + NFCI	1.74	1.24	2.47	2.02
AR + ADS + CFNAI + NFCI	1.21	1.14	1.58	1.17
AR + SPF median	1.27	1.11	–	1.22
AR + F^C	1.44	0.92	1.40	1.02
AR + F^C + F^L_{q-1} (Table 10)	0.90	0.83	0.78	0.82
AR + ADS + CFNAI + NFCI + F^C + F^L_{q-1}	1.13	1.24	1.22	1.14

Notes: Average MSFE ratio relative to the AR(1) bridge over the weeks shown. Every row is equation (9) with the regressors listed. ADS, CFNAI and NFCI use latest-vintage data, averaged over current-quarter observations treated as available under assumed publication lags of 7, 25 and 5 days, respectively (25 days after month-end for CFNAI); the last available observation is used if no current-quarter value is yet available; the SPF enters as the median forecast of annualized quarter-on-quarter real GDP growth for the target quarter, treated as available from week 7; for $h = 0$ in weeks 1–6 the previous survey's forecast of the same quarter is used, for $h = 1$ no forecast is available before week 7.

Table 12: Real GDP point forecasts before and since COVID

Forecasts made	2013–2019	2020–2026	2022–2026
$h = 0$	0.83 (7/9)	0.87 (0/9)	1.07 (0/9)
$h = 1$	0.82 (7/9)	0.89 (0/9)	1.07 (0/9)
$h = 2$	0.98 (2/9)	0.98 (0/9)	1.04 (0/9)
Quarters ($h = 0$)	28	26	18

Notes: MSFE of the two-factor bridge equation relative to the AR(1), averaged over weeks 5–13 of the quarter; in parentheses, the number of these weeks in which the bridge is significantly more accurate (Diebold–Mariano, $p < 0.10$). Same selection (1996–2012) and expanding window as Table 10; forecast-regression observations are excluded when either the target or lagged GDP falls in 2020Q1–2021Q4. Factor estimation retains these years, and pandemic quarters remain in the evaluation sample. The last target quarter is 2026Q2.

Post-2020 performance. Table 12 extends the evaluation through 2026Q2, retaining the measures and transformations selected through 2012. Factor models and forecast regressions continue to be re-estimated recursively. To limit the influence of extreme pandemic observations on the forecast regressions, estimation excludes observations whose target or lagged GDP falls in 2020Q1–2021Q4. This exclusion does not apply to factor estimation, and forecasts for pandemic quarters remain in the evaluation sample.

Averaged over weeks 5–13, MSFE ratios relative to the AR benchmark are 0.87, 0.89, and 0.98 at $h = 0, 1, 2$ for forecasts made from 2020 onward. None of the week-specific DM tests rejects equal accuracy in favor of the filing-based model at the 10 percent level. For forecasts made from 2022 onward, the corresponding ratios are 1.07, 1.07, and 1.04. The gains for current- and next-quarter growth documented before the pandemic therefore do not persist reliably in the later sample.

Shutdowns, reopening, and supply disruptions may have altered the relationship between firms' assessments and GDP growth. The results do not distinguish among these explanations, but show that recursive parameter estimation alone does not preserve the earlier gains when the selected measures remain fixed. They motivate reassessing variable selection and model specification as economic conditions change; Section 7.3 examines sensitivity to the transformation rule.

5.5 Density Forecasts and Downside Risk

Point forecasts describe expected growth, but firms' assessments may also contain information about the range of possible outcomes and the risk of a slowdown. Building on the quantile-regression approach of Adrian et al. (2019), I forecast seven conditional quantiles of first-release GDP growth. I retain the variable selection, weekly information sets, and 2013–2019 evaluation period of the point-forecast model, and assess accuracy both across the seven quantiles and with greater weight on the lower tail.

Conditional quantile estimation. For each forecast week w of quarter q and horizon $h \in \{0, 1, 2\}$, I estimate

$$Q_\tau(Y_{q+h} \mid \mathcal{I}_{q,w}) = \alpha_\tau + \rho_\tau Y_{q-L} + \beta_{C,\tau} F^C(q, w) + \beta_{L,\tau} F^L_{q-1}, \tag{11}$$

for $\tau \in \mathcal{T} = \{0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95\}$. The conditional quantile Q_τ is the growth threshold below which outcomes fall with probability τ, given the information $\mathcal{I}_{q,w}$ available at the forecast date. Thus, $\tau = 0.50$ describes the median outlook and $\tau = 0.05$ a downside outcome. The regressors are the latest published GDP growth observation Y_{q-L}, the current-quarter coincident input $F^C(q, w)$, and the previous-quarter leading input F^L_{q-1}.

The intercept α_τ and autoregressive coefficient ρ_τ vary across quantiles, as do the predictive coefficients $\beta_{C,\tau}$ and $\beta_{L,\tau}$. This allows a filing-based index to have different associations with weak and strong growth outcomes. Coefficients are also estimated separately by horizon and forecast week, with those subscripts suppressed.

Each regression minimizes the sum of quantile, or pinball, losses over the expanding sample of GDP outcomes published by the forecast date. For a realized outcome y and predicted quantile $\widehat{Q}_\tau$, the loss is

$$\ell_\tau(y - \widehat{Q}_\tau) = \big(\tau - \mathbf{1}\{y < \widehat{Q}_\tau\}\big)(y - \widehat{Q}_\tau), \tag{12}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. The asymmetric penalty makes each regression target its specified quantile. I sort the seven predicted values into increasing order when they cross, applying a discrete version of the rearrangement approach of Chernozhukov et al. (2010). Together, these quantiles describe the location and spread of the conditional distribution without fitting a parametric density.

Forecast evaluation. The benchmark is a quantile autoregression with an intercept and the latest published GDP observation. Alternative specifications replace or augment the filing-based indices with NFCI, using latest-vintage observations and an assumed five-day publication lag.

The continuous ranked probability score (CRPS) evaluates a predictive distribution by integrating twice the quantile loss over all probability levels. Lower scores indicate more accurate forecasts. With seven estimated quantiles, I use the discrete score

$$S(y) = \frac{2}{7}\sum_{\tau\in\mathcal{T}} \ell_\tau(y - \widehat{Q}_\tau), \qquad S_D(y) = \frac{2}{7}\sum_{\tau\in\mathcal{T}} (1-\tau)^2 \ell_\tau(y - \widehat{Q}_\tau). \tag{13}$$

The first score weights the seven quantiles equally; the second gives larger weights to lower quantiles to assess downside accuracy (Gneiting and Ranjan, 2011). Table 13 labels these finite-grid scores CRPS and left-tail CRPS, respectively.

For each horizon and forecast week, I divide the model's mean score by the benchmark's mean score and then average these ratios over the reported weeks. A ratio below one favors the model. Parentheses report the number of weeks in which a two-sided DM test rejects equal accuracy at 10 percent in favor of the model.

Forecast accuracy and downside risk. Averaged over weeks 5–13, the filing-based indices reduce the unweighted score by 18, 28, and 26 percent at $h = 0, 1, 2$; in weeks 1–4, the reductions are 29 percent for the current quarter and 35 percent one quarter ahead. At each reported week,

Table 13: GDP Quantile Forecast Accuracy and Downside Risk

	$h = 0$		$h = 1$		$h = 2$
Predictors (with the latest GDP)	weeks 1–4	weeks 5–13	weeks 1–4	weeks 5–13	weeks 5–13
CRPS, forecasts made 2013–2019					
Filing-based indices	0.71 (4/4)	0.82 (9/9)	0.65 (4/4)	0.72 (9/9)	0.74 (9/9)
NFCI	0.94 (4/4)	0.94 (9/9)	0.97 (0/4)	0.96 (0/9)	0.93 (1/9)
Filing-based indices + NFCI	0.77 (4/4)	0.84 (9/9)	0.73 (4/4)	0.77 (9/9)	0.75 (9/9)
Left-tail CRPS, forecasts made 2013–2019					
Filing-based indices	0.60 (4/4)	0.70 (9/9)	0.57 (4/4)	0.64 (9/9)	0.67 (9/9)
NFCI	0.70 (4/4)	0.81 (9/9)	0.70 (4/4)	0.73 (9/9)	0.63 (9/9)
Filing-based indices + NFCI	0.61 (4/4)	0.70 (9/9)	0.60 (4/4)	0.64 (9/9)	0.58 (9/9)
CRPS, validation: selection on 1996–2008, forecasts made 2009–2012					
Filing-based indices	0.80 (0/4)	1.02 (0/9)	0.87 (0/4)	0.93 (0/9)	0.93 (0/9)
NFCI	0.82 (0/4)	0.91 (0/9)	0.79 (0/4)	0.79 (0/9)	1.08 (0/9)
Filing-based indices + NFCI	0.74 (0/4)	0.92 (0/9)	0.78 (0/4)	0.83 (0/9)	1.06 (0/9)

Notes: Quantile regressions of first-release real GDP growth at $q + h$ on the latest published GDP figure and the predictors listed, estimated for $\tau \in \{0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95\}$ on an expanding window and re-estimated every week. Filing-based indices: the coincident flow factor $F^C(q, w)$ and the leading index F^L_{q-1} of the bridge model. CRPS denotes the discrete score computed as the mean of twice the pinball loss over the seven reported quantiles; it is not the continuous integral over all quantiles. Left-tail CRPS weights quantile τ by $(1 - \tau)^2$ (Gneiting and Ranjan, 2011). Each cell is the ratio to the quantile autoregression with the latest GDP only, averaged over the weeks shown; in parentheses, the number of weeks in which the Diebold–Mariano test on the loss difference rejects equal accuracy at 10% in favor of the model. Forecasts made before the selection date (1 February of the first forecast year) are excluded.

the two-sided DM test rejects equal accuracy at 10 percent in favor of the filing model. The earlier evaluation provides suggestive evidence of predictive value: the filing-based model lowers quantile forecast loss in four of five reported groups, although the improvements are not statistically significant.

The relative gains are larger under the downside-weighted score: 30, 36, and 33 percent in weeks 5–13 at $h = 0, 1, 2$. The NFCI alone improves this score more than the unweighted score. Adding it to the filing-based indices does not improve the unweighted score at any reported horizon, and provides little additional benefit under downside weighting at $h = 0, 1$. At $h = 2$, however, the combined model lowers the downside-weighted loss ratio from 0.67 to 0.58. This ranking suggests complementary information about longer-horizon downside risk.

Forecast intervals. Figure 11 compares one-quarter-ahead GDP forecast distributions formed in week 13 of each quarter over 2013–2019. Relative to the autoregressive quantile benchmark, conditioning on the filing-based indices produces narrower interquartile and 5th–95th percentile intervals, with broadly similar median forecasts. The difference is particularly pronounced in the lower tail: the filing-based forecasts assign less probability to negative growth during the expansion. The distributions also shift with economic conditions, weakening around the 2015–2016 industrial slowdown, strengthening during the subsequent acceleration, and moderating toward the end of the sample. These patterns suggest that the filing-based indices contribute information about the range

of plausible growth outcomes beyond that contained in lagged GDP.

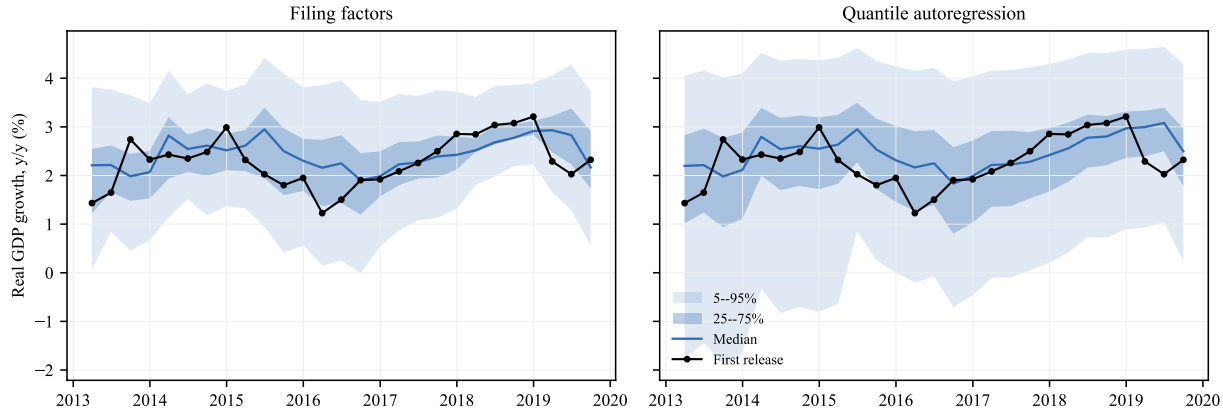


Figure 11: One-Quarter-Ahead GDP Forecast Distributions

Notes: Forecasts are made in week 13 of each quarter in 2013–2019 and plotted at the end of the forecast-origin quarter; the outcomes shown are for the following quarter. The left panel uses the two filing-based indices and the latest available GDP observation; the right panel reports the quantile autoregression. Bands denote the 5–95 and 25–75 percent intervals, the line denotes the conditional median, and dots denote first-release GDP.

Associations across quantiles. Figure 12 provides a complementary in-sample analysis of the relationship between filing-based indices and next-quarter GDP growth over 1997–2019. The quantile regressions include the current-quarter leading and coincident stock indices and control for current-quarter GDP growth. Unlike the real-time forecasting specification, it uses current-quarter values for all regressors. The coefficients measure conditional associations at different quantiles of GDP growth, expressed in percentage points per standard-deviation increase in each index.

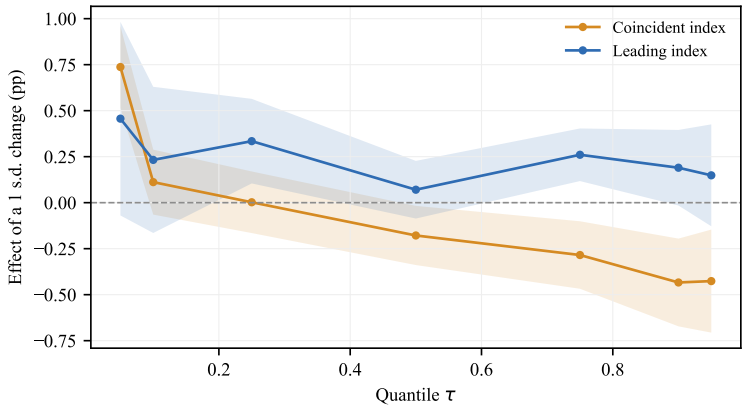


Figure 12: Filing-Based Indices and the Conditional Quantiles of GDP Growth

Notes: Quantile-regression coefficients on the standardized coincident and leading indices for real GDP growth one quarter ahead, 1997–2019, controlling for current GDP growth. Both predictors are descriptive stock indices, distinct from the real-time forecasting factors in equation (11). Shaded regions denote the reported pointwise 90 percent confidence intervals, without adjustment for serial correlation. Coefficients are conditional associations, measured in percentage points per standard deviation of the index.

The coincident coefficient is positive at the fifth percentile (0.74 percentage points per standard deviation), near zero at the 25th percentile, and negative at the median and upper quantiles. The leading coefficients are positive throughout, although the uncertainty bands are wide at several quantiles. The point estimates suggest that operating conditions are associated with the spread of subsequent growth outcomes, while the leading index has a broader positive association with growth.

6 Monthly Forecasts: Production, Inflation, and the Labor Market

I next examine whether filing-based measures improve monthly forecasts of production, inflation, and labor-market conditions. These outcomes capture different margins of firms' adjustment to demand and costs: output can respond through inventories and capacity utilization, while prices and employment reflect pricing and hiring decisions. A common forecasting design reveals where corporate soft information contributes most. The clearest gains are in industrial production point forecasts and year-ahead core-inflation quantile forecasts; labor-market gains are concentrated in unemployment forecasts at selected horizons.

6.1 Common Forecasting Design

The main evaluation covers 2013–2019. I select measures using data through 2012, hold their identities and transformations fixed, and produce forecasts on the last Friday of each month for the current month and three, six, and twelve months ahead. Models are estimated recursively using information available at each forecast date and compared with autoregressive benchmarks. An earlier evaluation repeats the selection through 2008 and forecasts 2009–2012. Only outcomes within each evaluation period are scored.

Targets and filing-based measures. The targets are first-release year-over-year growth in industrial production, core CPI, and nonfarm payrolls, and the 12-month change in the unemployment rate. Each is computed within the ALFRED vintage in which the target month first appears. Selection follows the GDP procedure, adapted to monthly data with three-month changes, a 72-month initial window, and a cross-correlation screen extending to 24 months. For each target, forward selection chooses a leading set S_L for the outcome three months ahead and a coincident set S_C for the current month. All baseline models use economy-wide weekly stock panels; Section 7.3 considers sector and flow-panel alternatives.

Table 14 reports the selected measures. Figure 13 plots their leading and coincident indices alongside the targets. The figure uses factor parameters estimated through 2019 to describe cyclical comovement. Forecasting instead re-estimates the factor models and standardization moments at the beginning of each quarter on an expanding sample starting in 1996. These estimates remain fixed within the quarter as new weekly observations enter the filter. Forecast regressions are re-estimated monthly using observations whose targets have been published.

Table 14: Selected Filing-Based Measures for Monthly Forecasts

Target	Leading set S_L (target 3 months ahead)	Coincident set S_C
Industrial production	Δ Sentiment, Debt concern (−), Δ Capacity constraints, Δ Uncertainty (−), Δ Wage growth expected	R&D direction, Inventory plan, Supplier delivery (−), Pricing power
Core CPI	Market share, Δ Revenue growth guidance (−), Δ Shareholder returns (−)	Rate sensitivity
Payroll employment	Δ Demand outlook, Debt concern (−), Δ Labor shortage, Hiring intentions, Sentiment	Δ Liquidity (−)
Unemployment rate	Δ Demand outlook (−), Debt concern, Δ R&D direction (−), Hiring intentions (−), Monetary-policy mentions	Wage growth expected (−)

Notes: Forward stepwise selection on 1996–2012 from the candidates passing the cross-correlation screen (leads of 0–24 months), with every variable offered in levels and as its 3-month change (Δ); monthly selection panel of all S&P 500 members' latest answers (economy-wide mean). (−) denotes a negative correlation with the selection target; unmarked measures have positive correlations.

Point forecasts. I compare a two-index bridge, an average of variable-specific mixed-data sampling (MIDAS) forecasts, and factor-MIDAS. Let t index calendar months, h denote the horizon in months, and Y_t be the first-release target. Each model is estimated separately by target and horizon. The bridge uses the coincident and leading factors extracted as in Section 5.3:

$$Y_{t+h} = \alpha + \rho Y_{t-L} + \beta_C F_t^C + \beta_L F_{t+\min(0,h-3)}^L + u_{t+h}, \qquad h \in \{0, 3, 6, 12\}. \tag{14}$$

The predictors F_t^C and F_t^L are monthly averages of the recursively estimated weekly coincident and leading factors for the target. They are distinct from the descriptive indices in Figure 13. The term Y_{t-L} is the latest published target, with L its publication lag in months at the forecast date. The coefficients α, ρ, β_C, and β_L are the intercept and predictive slopes, and u_{t+h} is the disturbance. The leading index enters at $t-3$ for the nowcast and at t otherwise, so the six- and twelve-month forecasts extend beyond its three-month selection horizon.

The MIDAS average estimates one regression per selected variable v on its last 26 weekly observations (Ghysels et al., 2007):

$$Y_{t+h} = \alpha_v + \rho_v Y_{t-L} + \sum_{\ell=0}^{25} w(\ell; \theta_v) x_{v,\tau_t-\ell} + u_{v,t+h}, \tag{15}$$

where v indexes selected filing series, including their chosen level or change transformation. The value $x_{v,\tau_t-\ell}$ is the standardized economy-wide weekly aggregate of series v, observed ℓ weeks before the forecast Friday τ_t. The Almon weights satisfy $w(\ell;\theta_v) = [\theta_{0v} + \theta_{1v}(\ell/26) + \theta_{2v}(\ell/26)^2]/26$, where θ_v contains three estimated coefficients. Each variable can therefore have a different lag profile and predictive slope. The intercept α_v, autoregressive coefficient ρ_v, and disturbance $u_{v,t+h}$ are specific



Figure 13: Filing-Based Indices and Monthly Economic Activity

Notes: Monthly averages of the weekly leading and coincident indices and the corresponding first-release target, standardized using 1996–2019 moments. Variables are selected on 1996–2012; factor parameters are estimated on 1996–2019 and the filter is extended through 2026. These are descriptive indices, distinct from the recursively estimated forecasting vintages. The unemployment target is the 12-month change in the rate; it and its indices are inverted. Shaded regions denote NBER recessions.

to each variable regression. Forecasts are averaged equally (Stock and Watson, 2004, Andreou et al., 2013). Factor-MIDAS applies the same weekly lag structure to the two indices (Marcellino and Schumacher, 2010). This comparison assesses whether forecasts benefit from retaining individual measures and their weekly histories.

Quantile forecasts and evaluation. Quantile regressions use the bridge predictors at the seven quantiles in Section 5.5, with the same recursive estimation and sorting of predicted quantiles. The benchmark is a quantile autoregression using the latest published target. I report the quantile-based CRPS and tail-weighted scores defined in that section. Tail weights are $(1-\tau)^2$ for production, payroll growth, and inflation, and τ^2 for the unemployment-rate change, emphasizing low growth, low inflation, and rising unemployment, respectively.

Tables 15 and 16 report loss ratios relative to the corresponding AR benchmarks; values below one indicate improvement. MSFE ratios follow equation (10), evaluated on common monthly forecast origins at each horizon. Two-sided DM tests use Newey–West bandwidth $\max(4, h)$ months. Stars

Table 15: Monthly Point Forecast Accuracy

	Main evaluation: 2013–2019				Earlier evaluation: 2009–2012			
Horizon (months)	0	3	6	12	0	3	6	12
Industrial production								
Bridge (two indices)	1.07***	0.97	1.02	1.03	0.82	0.39	0.38	0.62
MIDAS average	0.89***	0.74***	0.71**	0.91*	0.63*	0.45*	0.43	0.47**
Factor-MIDAS	0.85***	0.84	0.75*	0.89	0.59	0.34	0.33	0.66
Core CPI								
Bridge (two indices)	1.03**	1.02	0.98	0.89	0.96	0.96	0.81	0.34*
MIDAS average	0.99	0.94	0.89	0.82	1.00	0.83**	0.68**	0.55**
Factor-MIDAS	1.04	1.08***	1.00	0.88	0.99	0.83	0.71*	0.27**
Payroll employment								
Bridge (two indices)	1.20**	2.44**	2.11	1.23	0.77*	0.86	0.92	1.85***
MIDAS average	1.01	1.04	0.92	0.84	0.54*	0.52*	0.60	1.04
Factor-MIDAS	1.25	1.81*	1.63	1.00	0.51*	0.51	0.66	1.59*
Unemployment rate								
Bridge (two indices)	1.08*	1.48	1.66	1.57	0.73	0.63	0.58	0.73
MIDAS average	1.00	0.86	0.81	0.59**	0.79*	0.66*	0.69	1.05
Factor-MIDAS	0.99	1.06	1.29	1.44	0.69	0.46*	0.49	0.81

Notes: MSFE of each model divided by that of the autoregressive benchmark (the same equation without filing information); values below one favor the model. Forecasts are made on the last Friday of each month. Bridge: equation (14) with the coincident and leading indices built from all firms' latest answers. MIDAS: one MIDAS regression per selected variable on its last 26 weekly values, forecasts averaged with equal weights. Factor-MIDAS: the two weekly indices in a MIDAS regression with Almon lag polynomials. Validation: variables selected on 1996–2008, forecasts made 2009–2012. Diebold–Mariano test (Newey–West, $\max(4, h)$ lags) of equal accuracy, two-sided; stars mark significant differences in either direction, so a starred value above one means the model is significantly worse: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

indicate significant differences in either direction. Figure 14 illustrates point forecasts and quantile intervals at selected horizons. Its production, inflation, and payroll panels show bridge forecasts; the unemployment panel shows the MIDAS average, together with intervals from the bridge quantile model.

6.2 Industrial Production

Industrial production provides a monthly test of the real-activity information in corporate disclosures. Firms adjust output to expected sales, inventories, and available capacity. The selected leading measures capture changes in sentiment, capacity constraints, uncertainty, and expected wage growth, together with debt concern. The coincident measures cover R&D direction, inventory plans, supplier delivery, and pricing power. Together, they connect the demand outlook to the operating and financing conditions under which firms produce.

The MIDAS average delivers the most consistent point-forecast gains (Table 15). It reduces MSFE by 11, 26, 29, and 9 percent at horizons of zero, three, six, and twelve months in 2013–2019. The first three gains are significant at 5 percent or better and the last at 10 percent. In 2009–2012, its ratios range from 0.43 to 0.63, with significant gains at all horizons except six months. The bridge yields no significant improvement in either period and worsens the main-period nowcast.

Table 16: Monthly Quantile Forecast Accuracy

	Main evaluation: 2013–2019				Earlier evaluation: 2009–2012			
Horizon (months)	0	3	6	12	0	3	6	12
Industrial production								
CRPS	0.99	0.96	0.99	1.02	0.91	0.77	0.72	0.92
Left-tail CRPS	1.00	0.97	1.00	1.06^{***}	1.02	0.94	0.83	0.90
Core CPI								
CRPS	1.00	1.00	0.98	0.86^{***}	1.01	1.01	0.96	0.51^{**}
Left-tail CRPS	1.01	1.02^{**}	1.00	0.87^{***}	1.04	1.05	1.01	0.35^{**}
Payroll employment								
CRPS	1.02	1.16	1.23^{*}	1.23^{***}	0.91^{*}	1.01	1.13	1.50^{***}
Left-tail CRPS	1.03	1.18	1.30^{*}	1.28^{***}	0.98	0.97	0.97	1.06
Unemployment rate								
CRPS	1.04^{**}	1.05	1.01	1.37	0.92	0.97	0.99	0.74
Right-tail CRPS	1.05^{*}	1.06	1.14	1.93	0.88	0.83^{*}	0.87	0.51

Notes: Ratios of mean quantile forecast loss to that of a quantile autoregression using the latest published target; values below one favor the filing model. The filing model adds the bridge indices in equation (14). CRPS is twice the mean pinball loss across seven quantiles; tail scores additionally weight losses by $(1-\tau)^2$ for growth and inflation and τ^2 for the unemployment-rate change. Main evaluation: selection through 2012, forecasts made 2013–2019. Earlier evaluation: selection through 2008, forecasts made 2009–2012. Only outcomes within each evaluation period are scored. Two-sided Diebold–Mariano tests use Newey–West bandwidth $\max(4, h)$ months; stars indicate differences in either direction: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Factor-MIDAS improves the main-period nowcast at 1 percent and the six-month forecast at 10 percent; its remaining gains are not significant.

The largest MIDAS reductions occur three to six months ahead, consistent with production adjusting as firms revise sales expectations, inventories, and capacity use. Allowing each measure its own weekly lag profile accommodates differences in the timing of these adjustments, while forecast averaging combines information across operating conditions. The more consistent gains from the MIDAS average relative to the two-index models support retaining this detail when forecasting production. The bridge quantile model adds little to distributional accuracy in Table 16. Overall score ratios range from 0.96 to 1.02 in 2013–2019 and from 0.72 to 0.92 in 2009–2012, with no significant gains. Its twelve-month lower-tail score is significantly worse in the main period (1.06).

6.3 Core CPI Inflation

Core CPI inflation tests whether corporate assessments help anticipate price developments. The leading set comprises market share, changes in revenue guidance, and changes in shareholder-return priorities; the coincident measure is interest-rate sensitivity. These variables describe competitive conditions, expected sales, cash distribution, and exposure to financing costs, linking the forecast to firms' broader business conditions.

Point-forecast improvements are concentrated at twelve months. In 2013–2019, MSFE ratios at this horizon range from 0.82 to 0.89, although none is significant at 10 percent. In 2009–2012, the bridge, MIDAS average, and factor-MIDAS reduce twelve-month MSFE by 66, 45, and 73 percent,

respectively, with significance at 10, 5, and 5 percent. At shorter horizons, no model significantly improves on the AR benchmark in the main period. The year-ahead quantile forecasts show stronger and more consistent gains. The overall score falls by 14 percent in 2013–2019 and 49 percent in 2009–2012, while the lower-tail score falls by 13 and 65 percent. All four reductions are significant at 5 percent or better. Shorter-horizon quantile forecasts yield no significant gains.

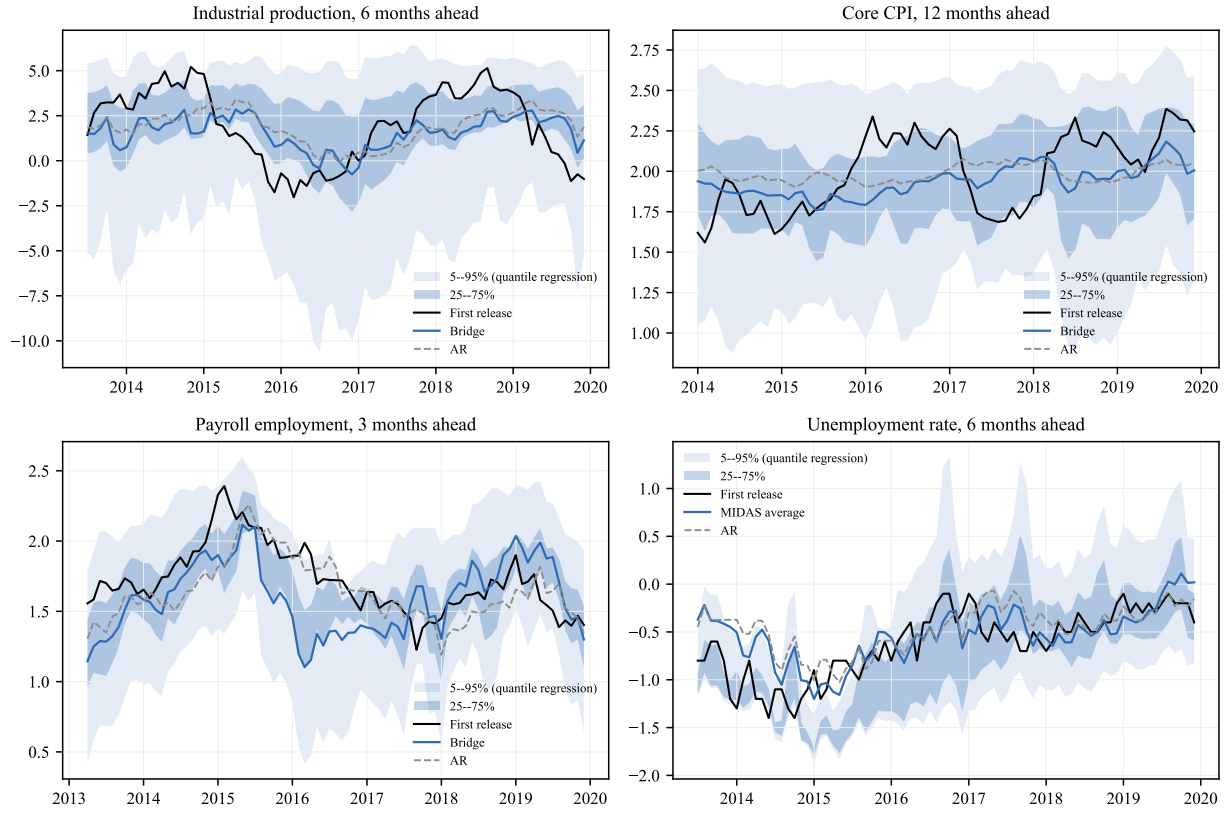


Figure 14: Monthly Forecasts and Prediction Intervals

Notes: Forecasts made in 2013–2019 are plotted at the target month. Industrial production, core CPI, and payrolls are measured as year-over-year percentage growth; unemployment is the 12-month change in the rate, in percentage points. The panels show industrial production six months ahead, core CPI twelve months ahead, payroll employment three months ahead, and the unemployment-rate change six months ahead. Black: first release; gray dashed: AR benchmark. Blue: bridge forecast for industrial production, core CPI, and payrolls; MIDAS average for unemployment. Light and dark shading denote the 5–95 and 25–75 percent intervals from quantile regressions on the two bridge indices and the latest published target. The shaded intervals and the unemployment point forecast therefore come from different models.

The concentration of improvements at twelve months is consistent with business conditions passing gradually into prices as firms adjust spending and pricing decisions. Market-share assessments convey competitive pressure, while interest-rate sensitivity reflects exposure to financing costs and interest-sensitive demand. The quantile results indicate that these measures help characterize the distribution of subsequent inflation, including its lower tail. This information extends beyond direct reports of input costs: the selected measures chiefly describe the conditions under which firms make

pricing decisions. Appendix H reports complementary associations with core-inflation outcomes and professional forecast errors.

6.4 The Labor Market

Payroll growth and the change in unemployment capture different aspects of labor-market adjustment. Payrolls measure net employment creation; unemployment also reflects job finding and labor-force participation. Corporate hiring intentions and demand assessments can inform both outcomes, but their aggregate implications depend on how firms adjust labor input.

Payroll employment. The leading set contains changes in demand and labor-shortage risk, hiring intentions, sentiment, and debt concern; the coincident measure is the change in liquidity. In 2013–2019, no point-forecast specification significantly improves on the AR benchmark. The MIDAS average lowers twelve-month MSFE by 16 percent, but its short-horizon ratios are near one. In 2009–2012, its nowcast and three-month forecast improve significantly, whereas its twelve-month ratio exceeds one. The bridge quantile model also performs poorly in the main period: both overall and lower-tail scores are significantly worse at six and twelve months. Thus, the selected hiring and demand measures do not deliver consistent payroll forecast gains across periods.

Unemployment. The leading set combines changes in demand and R&D direction, hiring intentions, debt concern, and monetary-policy mentions; expected wage growth is coincident. Demand, R&D, hiring, and wages have negative selection signs, consistent with stronger business conditions accompanying lower unemployment. Debt concern and monetary-policy mentions have positive signs.

The MIDAS average reduces MSFE by 14 and 19 percent at three and six months in 2013–2019, neither significantly, and by 41 percent at twelve months, significant at 5 percent. The earlier evaluation instead shows significant gains for the nowcast and three-month forecast, while the twelve-month ratio is 1.05. Neither the bridge nor factor-MIDAS significantly improves point forecasts in the main period. Quantile gains are also concentrated in the earlier period: the three-month upper-tail score falls by 17 percent, significant at 10 percent, while no quantile gain is significant in 2013–2019.

The contrast with industrial production is consistent with differences in firms' adjustment margins. Firms can change hours, capacity utilization, and inventories before changing headcounts. Aggregate payroll growth also reflects separations, and unemployment reflects participation and job finding. These additional margins help explain why operating assessments can be more informative about production than employment. Across the labor-market targets, forecast averaging yields the clearest gains for unemployment, with the useful horizon varying across evaluation periods.

Table 17: Monthly Point Forecast Accuracy After 2020

		Forecasts made 2020–2026				2022–2026			
Target	Model	$h = 0$	$h = 3$	$h = 6$	$h = 12$	$h = 0$	$h = 3$	$h = 6$	$h = 12$
Industrial production	Bridge	1.01	1.07	0.89	1.09	1.12	1.57	1.91	1.01
	MIDAS average	1.00	1.05	1.32	1.68	1.03	1.52	2.69	1.41
	Factor-MIDAS	0.98	0.96	0.93	1.20	0.95	1.40	1.82	1.49^{**}
Core CPI	Bridge	0.98	1.00	1.08	1.14^{**}	0.98	0.99	1.18	1.37
	MIDAS average	1.03	1.06^{*}	1.11^{*}	1.07	1.01	1.05	1.11	1.12
	Factor-MIDAS	0.99	1.01	1.15^{**}	1.20^{*}	1.00	0.99	1.16	1.22
Unemployment rate	Bridge	0.99^{*}	0.95^{**}	0.93	1.28^{*}	0.91	0.87	1.19	5.20^{***}
	MIDAS average	0.98	0.97	0.96	1.21	0.86	0.60^{*}	0.50^{*}	0.30^{*}
	Factor-MIDAS	0.96	0.90	0.91	1.32^{*}	1.01	1.43	1.86	5.49^{**}

Notes: MSFE relative to the autoregressive benchmark for forecasts made on the last Friday of each month; values below one favor the filing model. Horizons h are in months. Selection and models follow Table 15, with measures selected through 2012. Forecast regressions exclude observations whose dependent variable or lagged target falls in March 2020–December 2021; factor estimation uses the expanding historical sample. The last target month is August 2026. Two-sided Diebold–Mariano tests use Newey–West bandwidth $\max(4, h)$ months; stars indicate differences in either direction: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

6.5 Stability After 2020

Table 17 extends the point forecasts for industrial production, core CPI, and unemployment through August 2026. Selection remains fixed at its 2012 cutoff. Forecast regressions omit training observations whose dependent variable or lagged target falls in March 2020–December 2021; factor estimation continues to use the expanding historical sample.

The pre-pandemic production and inflation gains do not persist. No production model significantly improves on the AR benchmark over 2020–2026. In 2022–2026, the MIDAS average is worse at every horizon, with ratios of 1.52 and 2.69 at three and six months, where its earlier gains were largest. For core inflation, every twelve-month ratio exceeds one in both later windows; the bridge and factor-MIDAS deteriorations are significant over 2020–2026. Figure 13 also shows the sharp post-2020 increase in inflation relative to the movements in its selected indices.

Unemployment follows a different pattern. Over 2020–2026, the bridge reduces nowcast and three-month MSFE by 1 and 5 percent, respectively, both significantly. In 2022–2026, the MIDAS average lowers three-, six-, and twelve-month MSFE by 40, 50, and 70 percent, each significant at 10 percent. The bridge and factor-MIDAS, however, perform substantially worse at twelve months, with ratios of 5.20 and 5.49. These comparisons reinforce the value of retaining individual measures for unemployment forecasting and show that the contribution of corporate soft information varies across targets, horizons, and economic conditions.

7 Robustness and Look-Ahead Risk

I assess two sources of look-ahead risk and the sensitivity of the forecasts to specification choices. A truncation audit checks whether later-dated inputs affect historical forecasts. Answer-treatment

comparisons examine whether forecast gains depend on contextual inferences by the LLM, whose training may include information published after the filing date. I then vary the transformation rules, panel construction, and use of weekly predictor histories.

7.1 Forecast Information Sets

The truncation audit removes inputs dated after each forecast origin and rebuilds the forecasting pipeline. It covers 17 target–date cases across eight distinct dates: five GDP cases and three cases for each monthly target. Removed inputs include later filings, S&P 500 membership records, ALFRED vintages, and NFCI observations unavailable under the assumed publication lag. Future end dates of ALFRED vintages are also removed. Using the retained inputs, I reconstruct the panels, winsorization bounds, factors, and forecast regressions and compare the resulting forecasts with the full run.

Table 18 shows that the forecasts are unchanged to numerical tolerance. The largest point-forecast difference is 7.4×10^{-9} percentage points, and the largest quantile difference is 1.6×10^{-14}. The larger point differences arise from the numerical optimizer used for exponential Almon weights. As a diagnostic comparison, an earlier implementation using within-quarter winsorization bounds produces differences of order 10^{-3}. The final implementation uses trailing winsorization bounds, prioritizes EDGAR dates over extracted dates, and scores forecasts only after the selection data have been published. The audit verifies invariance to later-dated inputs in the cases examined; the next comparison addresses the content of the extracted answers.

7.2 Explicit Disclosures and Inferred Answers

Dating an answer by its filing date establishes when the source became public, but an LLM's contextual inference may also draw on knowledge acquired from later material. I examine the role of inferred answers by repeating selection and forecasting under two alternative treatments. The *explicit-only* treatment retains an answer only when the model classifies it as explicitly stated in the firm's latest filing. The *explicit net-share* treatment replaces inferred and missing answers with a fixed, variable-specific training-sample median of inferred answers. Under the latter treatment, variation around that reference value reflects the content and prevalence of explicit disclosures. Each treatment reselects measures using data through 2012 and evaluates forecasts in 2013–2019.

Table 19 compares these treatments within sector and economy-wide panels.[8] For one-quarter-ahead GDP forecasts with sector panels, the explicit-only and net-share MSFE ratios are 0.72 and 0.62, compared with 0.73 using all answers in this comparison. With economy-wide panels, the corresponding ratios are 0.78 and 0.64, against 0.53 using all answers. Thus, explicit-only forecasts improve on the AR benchmark under both panel constructions. Restricting the answers improves accuracy in the sector specification but reduces it in the economy-wide specification.

[8]The GDP answer-treatment comparison reselects both sets and excludes the selected leading series from the coincident candidates. Unlike the main selection procedure, it can retain another transformation of the same measure in the coincident set. Its all-answer column is therefore the reference for comparisons within this table.

Table 18: Forecast Invariance to Future-Input Removal

Target	Forecast date	Records	Max. difference, point	Max. difference, quantiles
Real GDP	2013-04-19	15	2.1e-14	1.4e-14
Real GDP	2014-10-03	15	1.8e-09	1.6e-14
Real GDP	2016-02-12	15	6.4e-09	9.3e-15
Real GDP	2017-09-29	15	4.9e-10	6.2e-15
Real GDP	2019-05-03	15	7.4e-09	1.3e-14
Industrial production	2013-05-31	20	2.2e-16	2.2e-16
Industrial production	2016-02-26	20	2.2e-16	1.8e-15
Industrial production	2018-11-30	20	4.4e-16	4.4e-16
Core CPI	2013-05-31	20	2.2e-16	4.4e-16
Core CPI	2016-02-26	20	2.2e-16	4.4e-16
Core CPI	2018-11-30	20	4.4e-16	4.4e-16
Payroll employment	2013-05-31	20	2.2e-16	2.2e-16
Payroll employment	2016-02-26	20	2.2e-16	4.4e-16
Payroll employment	2018-11-30	20	2.2e-16	2.2e-16
Unemployment rate	2013-05-31	20	2.2e-16	2.2e-16
Unemployment rate	2016-02-26	20	2.2e-16	2.2e-16
Unemployment rate	2018-11-30	20	8.3e-17	2.2e-16

Notes: The audit removes inputs dated after each forecast origin: filings by EDGAR date, S&P 500 membership weeks, ALFRED vintage start dates and future vintage end dates, and NFCI observations under the assumed five-day publication lag. Panels, winsorization bounds, factors, and regressions are rebuilt. Differences are maximum absolute deviations from the full-run forecasts across the audited models and horizons, in percentage points. Records counts the matched point-forecast rows and model–horizon quantile rows; each quantile row contains seven forecasts. The inverse-MSFE-weighted MIDAS combination is excluded because its weights require the sequence of earlier forecasts. Point differences near 10^{-8} reflect numerical optimization of exponential Almon weights.

For six-month industrial production bridge forecasts, retaining all answers gives lower MSFE than either restricted treatment. The explicit-only ratios are 1.00 with sector panels and 1.07 with economy-wide panels, compared with 0.93 and 1.02 using all answers. This comparison evaluates the monthly bridge; the MIDAS average is the source of the main production gains in Section 6.2.

The central finding is that the GDP gains can be obtained from answers classified as explicit disclosures. Firm-level evidence in Table 5 also shows that inferred demand assessments, production plans, and hiring intentions predict subsequent firm outcomes. These tests establish the predictive content of different answer treatments; because the same LLM extracts and classifies both types of answer, they do not separately identify the contribution of knowledge learned after the filing date.

7.3 Specification Choices

Levels and changes. Table 20 compares six transformation rules, repeating selection and forecasting under each rule. Offering both levels and changes, as in the main specification, gives MSFE ratios of 0.83, 0.82, and 0.98 at $h = 0, 1, 2$ over weeks 5–13. The KPSS rule produces ratios of 0.80, 0.70, and 0.89. Both rules select leading sets dominated by changes in firms' outlooks and plans. These changes capture revisions to expected activity, while levels describe prevailing conditions.

The ADF rules without a trend improve current- and next-quarter forecasts but worsen forecasts two quarters ahead. Keeping every measure in levels, or using ADF with a trend, produces ratios above one at all three horizons. Transformation choices also alter the selected leading measures.

Table 19: Forecast Accuracy by Answer Treatment

Target	Panel	All answers	Explicit only	Explicit net share
Real GDP, next quarter	Sector	0.73	0.72	0.62
	Economy-wide	0.53	0.78	0.64
Industrial production, 6 months	Sector	0.93	1.00	1.06
	Economy-wide	1.02	1.07	1.33

Notes: MSFE relative to an autoregressive benchmark for forecasts made in 2013–2019; values below one favor the filing model. Measures are reselected through 2012 for each answer treatment. Explicit only: inferred answers are missing, and the latest filing must state the answer explicitly. Explicit net share: inferred and missing answers are replaced by the variable's training-sample median inferred value, so variation around this reference reflects explicit disclosures. Answers are carried forward for at most 52 weeks, with no minimum answer count. Panels use 11 sector means or economy-wide means. GDP: the flow-coincident and lagged-stock-leading bridge, one quarter ahead, with ratios averaged over weeks 5–13. Industrial production: the stock-index bridge, six months ahead. Comparisons are within the reselected configurations shown here.

For example, the 5-percent ADF, trend-ADF, and KPSS rules all yield an average ratio of 0.66 in 2009–2012, yet their one-quarter-ahead ratios in 2013–2019 range from 0.70 to 3.84. The strongest later-period results come from rules that retain revisions to revenue guidance, inventories, hiring, and wages.

Table 20: GDP Forecast Accuracy under Alternative Transformations

Rule (1996–2012)	Leading set S_L	Validation	Test 2013–19, weeks 5–13		
		2009–12	$h = 0$	$h = 1$	$h = 2$
No test, all levels	Supply-chain risk, Fiscal mentions	0.85	1.42	2.76	4.56
ADF, 5%	Pricing direction (Δ), Supplier delivery, R&D direction (Δ), Debt concern	0.66	0.83	0.87	1.47
ADF, 10%	Pricing direction (Δ), Supplier delivery, R&D direction (Δ), Debt concern	0.72	0.83	0.87	1.47
ADF with trend, 5%	Sentiment (Δ), Innovation, Labor shortage (Δ), Capacity constraints	0.66	1.93	3.84	5.48
KPSS, 5%	Revenue growth guidance (Δ), Inventory vs demand (Δ), Hiring intentions (Δ), Wage growth expected (Δ), Forward visibility (Δ)	0.66	0.80	0.70	0.89
No test, level and change both offered	Revenue growth guidance (Δ), Supply-chain risk, Inventory vs demand (Δ), Wage growth expected (Δ), Hiring intentions (Δ)	0.72	0.83	0.82	0.98

Notes: Each rule repeats variable selection and forecasting. ADF denotes the augmented Dickey–Fuller test; KPSS denotes the Kwiatkowski–Phillips–Schmidt–Shin test. Δ denotes a quarterly change. The leading sets shown use selection through 2012; the coincident set is unchanged except under trend-ADF and KPSS. Entries are mean MSFE ratios relative to the AR benchmark. Earlier evaluation: selection through 2008, forecasts made 2009–2012, averaged over $h = 0, 1$ and all weeks. Main evaluation: selection through 2012, forecasts made 2013–2019, averaged over weeks 5–13 at each horizon. Values below one favor the filing model.

Monthly panel construction. Preserving sector detail helps selected monthly bridge forecasts. For industrial production, three- and six-month ratios fall to 0.92 and 0.93, from 0.97 and 1.02 with economy-wide panels. The twelve-month unemployment ratio falls from 1.57 to 0.64. In contrast, the twelve-month core-inflation ratio rises from 0.89 to 1.01, and payroll ratios remain above one

at three to twelve months. Sector detail is therefore useful for some targets, while economy-wide aggregation remains the common baseline for the monthly analysis.

Replacing the coincident stock factor with a flow factor changes which filings enter the monthly bridge. Stock panels retain firms' latest assessments, whereas flow panels emphasize newly disclosed information. For industrial production, the economy-wide flow specification gives ratios of 0.98 and 1.05 at three and six months, close to the stock results. The twelve-month unemployment ratio improves to 0.68, while the corresponding core-inflation ratio rises to 1.01 and payroll ratios increase at every horizon. The balance between retaining earlier assessments and emphasizing new disclosures varies across targets.

Weekly lag specifications for GDP. I also replace the bridge's predictor averages with factor-MIDAS models using Almon, exponential Almon, or unrestricted weights (Foroni et al., 2015). These models have higher MSFE than the bridge in 2013–2019, although they perform more favorably in 2009–2012. An equal-weight average of variable-specific MIDAS forecasts is more competitive. In weeks 5–13, its current- and next-quarter ratios are 0.82 and 0.79, compared with 0.83 and 0.82 for the bridge. In weeks 1–4, it improves the nowcast (0.74 versus 0.90) but raises next-quarter MSFE (0.87 versus 0.78). The bridge captures much of the predictive information with two indices, while variable-specific histories provide additional gains at selected horizons and stages of the quarter.

8 Conclusion

Corporate financial reports are far more than repositories of historical accounting numbers; their narratives contain rich, forward-looking soft information about the trajectory of the aggregate economy. In this paper, I develop a structured LLM-based extraction framework to translate 60,236 Form 10-K and 10-Q filings by S&P 500 firms between 1996 and 2026 into 38 filing-based measures of firm-level demand, operations, supply chains, costs, risks, and expectations. By dating every observation strictly by its public EDGAR filing date and recording verbatim textual evidence, confidence ratings, and explicit-versus-implied disclosure status, this approach turns unstructured regulatory prose into a transparent, real-time panel of three million economic assessments.

The empirical findings confirm that these filing-based measures are economically meaningful. At the micro level, 28 of the 38 measures significantly predict reporting firms' subsequent accounting outcomes. At the macro level, their weekly aggregates track established macroeconomic benchmarks and business-cycle fluctuations across real activity, inventories, labor markets, supply chains, and inflation.

Crucially, this soft information translates into meaningful real-time forecasting gains. Selecting variables on 1996–2012 and extracting weekly leading and coincident factors via dynamic factor models, I find that corporate narratives reduce the mean squared forecast error of a real-time autoregressive benchmark for first-release real GDP growth by 17% in the current quarter and 18% one quarter ahead during the 2013–2019 evaluation period. In quantile regressions, the same filing-

based indices improve density forecasts of GDP growth by 18–28% and downside tail-risk forecasts by 30–36%. Extending the analysis to monthly targets shows that filing narratives also improve point forecasts of industrial production across horizons and year-ahead density forecasts of core CPI inflation, even though they do not yield consistent out-of-sample gains for payroll employment. Moreover, the GDP gains are concentrated in first-release estimates, indicating that timely corporate narratives anticipate the early source data underlying initial national accounts releases.

The results also identify where further work is needed. The weaker performance after 2020 calls for examining how the relationship between corporate assessments and aggregate outcomes changes across economic conditions. Sensitivity to variable transformations and the possibility of hindsight in LLM extraction warrant further evaluation. The central finding is that routine corporate disclosures contain useful information about firms' expectations and constraints that can inform assessments of current activity and forecasts of future growth.

References

Abdalla, A. M., Carabias, J. M., and Patatoukas, P. N. (2021). The real-time macro content of corporate financial reports: A dynamic factor model approach. *Journal of Monetary Economics*, 118:260–280.

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109(4):1263–1289.

Ahir, H., Bloom, N., and Furceri, D. (2022). The world uncertainty index. Technical report, National bureau of economic research.

Albrizio, S., Dizioli, A., and Simon, P. V. (2023). Mining the gap: Extracting firms' inflation expectations from earnings calls. IMF Working Paper 2023/202, International Monetary Fund.

Andreou, E., Ghysels, E., and Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, 31(2):240–251.

Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

Audrino, F., Maly, J., and Stalder, S. (2024). Quantifying uncertainty: a new era of measurement through large language models.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4):1593–1636.

Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. *Handbook of Economic Forecasting*, 2:195–237.

Bańbura, M., Giannone, D., and Reichlin, L. (2011). Nowcasting. In Clements, M. P. and Hendry, D. F., editors, *The Oxford Handbook of Economic Forecasting*, chapter 7. Oxford University Press, Oxford.

Bańbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of applied econometrics*, 29(1):133–160.

Benigno, G., di Giovanni, J., Groen, J. J. J., and Noble, A. I. (2022). The GSCPI: A new barometer of global supply chain pressures. Staff Report 1017, Federal Reserve Bank of New York.

Bernanke, B. S., Gertler, M., and Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. *Handbook of macroeconomics*, 1:1341–1393.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Blinder, A. S., Lovell, M. C., and Summers, L. H. (1981). Retail inventory behavior and business

fluctuations. *Brookings Papers on Economic Activity*, 1981(2):443–520.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3):623–685.

Bloom, N., Hassan, T. A., Kalyani, A., Lerner, J., and Tahoun, A. (2021). The diffusion of disruptive technologies.

Board of Governors of the Federal Reserve System (2021). Industrial production and capacity utilization. G.17 Statistical Release. Release date: July 15, 2021.

Bonsall IV, S. B., Bozanic, Z., and Fischer, P. E. (2013). What do management earnings forecasts convey about the macroeconomy? *Journal of Accounting Research*, 51(2):225–266.

Buehlmaier, M. M. and Whited, T. M. (2018). Are financial constraints priced? evidence from textual analysis. *The Review of Financial Studies*, 31(7):2693–2728.

Bybee, J. L. (2023). The ghost in the machine: Generating beliefs with large language models. *arXiv preprint arXiv:2305.02823*, pages 349–389.

Bybee, L., Kelly, B., Manela, A., and Xiu, D. (2024). Business news and business cycles. *The Journal of Finance*, 79(5):3105–3147.

Caldara, D. and Iacoviello, M. (2022). Measuring geopolitical risk. *American economic review*, 112(4):1194–1225.

Caldara, D., Iacoviello, M., Molligo, P., Prestipino, A., and Raffo, A. (2020). The economic effects of trade policy uncertainty. *Journal of Monetary Economics*, 109:38–59.

Carriero, A., Pettenuzzo, D., and Shekhar, S. (2024). Macroeconomic forecasting with large language models. *arXiv preprint arXiv:2407.00890*.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

Clayton, C., Coppola, A., Maggiori, M., and Schreger, J. (2025). Geoeconomic pressure. Technical report, National Bureau of Economic Research.

Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–2678.

Cortes, G. S., Sethuraman, M., and Silva, F. B. G. (2023). Monetary policy and corporate communication: Evidence from conference calls. SSRN Working Paper.

Davis, A. K., Piger, J. M., and Sedor, L. M. (2012). Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research*, 29(3):845–868.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North*

American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), pages 4171–4186.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Druz, M., Petzev, I., Wagner, A. F., and Zeckhauser, R. J. (2020). When managers change their tone, analysts and investors change their tune. *Financial Analysts Journal*, 76(2):47–69.

Estrella, A. and Mishkin, F. S. (1998). Predicting U.S. recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, 80(1):45–61.

Faria-e Castro, M. and Leibovici, F. (2024). Artificial intelligence and inflation forecasts. *Federal Reserve Bank of St. Louis Review*, 106(12):1–14.

Foroni, C., Marcellino, M., and Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A*, 178(1):57–82.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772.

Gennaioli, N., Ma, Y., and Shleifer, A. (2016). Expectations and investment. *NBER macroeconomics annual*, 30(1):379–431.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of monetary economics*, 55(4):665–676.

Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422.

Gosselin, M.-A. and Taskin, T. (2023). What can earnings calls tell us about the output gap and inflation in canada? Technical report, Bank of Canada Staff Discussion Paper.

Hansen, A. L., Horton, J. J., Kazinnik, S., Puzzello, D., and Zarifhonarvar, A. (2026). Simulating the survey of professional forecasters. *Available at SSRN 5066286*.

Hansen, A. L. and Kazinnik, S. (2024). Can chatgpt decipher fedspeak? *Available at SSRN 4399406*.

Hansen, S. and McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99:S114–S133.

Hassan, T. A., Hollander, S., van Lent, L., and Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, 134(4):2135–2202.

Hoberg, G. and Maksimovic, V. (2015). Redefining financial constraints: A text-based analysis. *The Review of Financial Studies*, 28(5):1312–1352.

Hoberg, G. and Phillips, G. (2016). Text-based network industries and endogenous product differ-

entiation. *Journal of Political Economy*, 124(5):1423–1465.

Isler, C. (2025). Words matter: Forecasting economic downside risks with corporate textual data. Technical Report 2511.04935, arXiv.

Jha, M., Qian, J., Weber, M., and Yang, B. (2024). Chatgpt and corporate policies. Technical report, National Bureau of Economic Research.

Jiang, F., Lee, J., Martin, X., and Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, 132(1):126–149.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182.

Kalyani, A. and Ozkan, S. (2025). Theory meets textual analysis: Measuring firm-level labor cost pressures and inflation pass-through. Working Paper 2025-021, Federal Reserve Bank of St. Louis. Revised January 30, 2026.

Khan, A. and Thomas, J. K. (2007). Inventories and the business cycle: An equilibrium analysis of (s, s) policies. *American Economic Review*, 97(4):1165–1188.

Konchitchki, Y. and Patatoukas, P. N. (2014). Accounting earnings and gross domestic product. *Journal of Accounting and Economics*, 57(1):76–88.

Lansford, B., Lev, B., and Wu Tucker, J. (2013). Causes and consequences of disaggregating earnings guidance. *Journal of Business Finance & Accounting*, 40(1-2):26–54.

Lopez-Lira, A. and Tang, Y. (2023). Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619*.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Marcellino, M. and Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 72(4):518–550.

Mayew, W. J., Pinto, J., and Wu, X. (2026). On the usefulness of guidance reports: Wj mayew. *Review of Accounting Studies*, pages 1–63.

Merkley, K. J. (2014). Narrative disclosure and earnings performance: Evidence from r&d disclosures. *The Accounting Review*, 89(2):725–757.

Nallareddy, S. and Ogneva, M. (2017). Predicting restatements in macroeconomic indicators using accounting information. *The Accounting Review*, 92(2):151–182.

Pettenuzzo, D., Sabbatucci, R., and Timmermann, A. (2024). Financial statements and macroeconomic dynamics. *Available at SSRN 4922676*.

Phan, M. T. (2025). Forecasting the macroeconomy with corporate disclosures and lan-

guage models (october 12, 2025). *Available at SSRN: https://ssrn.com/abstract=5594411 or http://dx.doi.org/10.2139/ssrn.5594411*.

Shivakumar, L. and Urcan, O. (2017). Why does aggregate earnings growth reflect information about future inflation? *The Accounting Review*, 92(6):247–276.

Soto, P. E. (2025). Research in commotion: Measuring AI research and development through conference call transcripts. Finance and Economics Discussion Series 2025-011, Board of Governors of the Federal Reserve System.

S&P Dow Jones Indices (2026). S&P 500®. https://www.spglobal.com/spdji/en/indices/equity/sp-500/. Accessed October 20, 2026.

Stock, J. H. and Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41(3):788–829.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168.

Ulrich Ruch, F. and Taskin, T. (2024). Global demand and supply sentiment: Evidence from earnings calls. *Oxford Bulletin of Economics and Statistics*, 86(2):314–334.

U.S. Bureau of Economic Analysis (2021a). Gross domestic product, second quarter 2021 (advance estimate) and annual update. News Release BEA 21-36. Release date: July 29, 2021.

U.S. Bureau of Economic Analysis (2021b). Gross domestic product, (third estimate), gdp by industry, and corporate profits (revised), 2nd quarter 2021. News Release BEA 21-47. Release date: September 30, 2021.

U.S. Securities and Exchange Commission (2021). How to read a 10-k/10-q. Accessed: 2026-05-24.

U.S. Securities and Exchange Commission (2023). *Regulation S-K, 17 C.F.R. § 229.101 et seq.* Available at: https://www.ecfr.gov/current/title-17/chapter-II/part-229.

U.S. Securities and Exchange Commission (2024). Approved artificial intelligence disclosure recommendation. Technical report, U.S. Securities and Exchange Commission.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Windsor, C. and Zang, M. (2023). Firms' price-setting behaviour: Insights from earnings calls. Research Discussion Paper 2023-06, Reserve Bank of Australia.

Zarifhonarvar, A. (2025). Generating inflation expectations with large language models. *Journal of Monetary Economics*, page 103859.

Appendices

A Sector Coverage

Table A.1: Sector distribution of filings

Sector	Filings	%	Firms
Consumer Discretionary	9,138	15.2	173
Financials	8,553	14.2	164
Technology	8,325	13.8	176
Industrials	7,952	13.2	148
Healthcare	6,772	11.2	131
Consumer Staples	4,469	7.4	70
Materials	4,002	6.6	78
Utilities	3,792	6.3	56
Energy	3,759	6.2	82
Real Estate	2,106	3.5	43
Telecommunications	1,315	2.2	37
Other	53	0.1	1
Total	60,236	100.0	1,159

Notes: 10-Q and 10-K filings by S&P 500 members, January 1996–September 2026. Each firm is assigned the sector it is given most often across its filings (the questionnaire's sector question). Other: News Corporation, whose filings are most often given no specific sector.

B 10-K/10-Q Filing Preprocessing

Since 10-Q and 10-K filings are lengthy documents, I preprocess each filing to reduce noise and token cost before extraction. Keeping irrelevant sections (e.g., corporate governance, executive compensation) would add no informational value to our questionnaire while increasing processing cost. Similarly, formatting artifacts inherited from the original HTML source can distract the language model. I address this in two steps: selecting relevant sections and cleaning the raw text.

1. **Data collection.**
   (a) Identify every company that was a member of the S&P 500 at any time between January 1996 and September 2026 from the weekly index constituent lists, and record each spell of membership.
   (b) Map each membership spell to the SEC identifier (CIK) of the company that filed during it. Tickers are reused and passed between companies (for example, after mergers), so the mapping uses CRSP security histories and EDGAR's record of former company names rather than today's ticker list.
   (c) From SEC EDGAR, download every 10-K (including the 10-K405 form used until 2003) and 10-Q filed while the company was a member, and retain only the main document and its metadata (ticker, CIK, accession number, filing date), discarding supplementary attachments (e.g., contracts, officer certifications).
2. **Section selection.**
   (a) Read through a subset of filings across different sectors and time periods to determine which sections contain information relevant to our 50 variables.
   (b) For 10-K filings, retain: Item 1 (Business Description), Item 1A (Risk Factors), Item 5 (Market for Common Equity), Item 7 (MD&A), Item 7A (Market Risk Disclosures), Item 8 (Financial Statements).
   (c) For 10-Q filings, retain: Part I Items 1–3 (Financial Statements, MD&A, Market Risk) and Part II Items 1–1A (Legal Proceedings, Risk Factors).
   (d) Exclude all remaining sections: corporate governance, executive compensation, mine safety disclosures, internal control procedures, and signature pages. These sections offer limited informational value relative to their token cost.
   (e) Filings before about 2001 are plain text rather than HTML; the same items are located by their headings, and a short 10-Q whose items cannot be separated is kept whole.
3. **De-duplication.**
   (a) Remove duplicate text that appears in more than one section.[9]
4. **Text cleaning.**
   (a) Replace non-standard spaces (non-breaking, zero-width, thin) with regular spaces.
   (b) Replace curly (smart) quotes with straight quotes.
   (c) Replace en-dashes and em-dashes with regular hyphens.

[9] The EDGAR document structure sometimes creates overlapping fragments, especially from financial tables (e.g., the same line items appear under both the full table section and as separate titled subsections).

(d) Collapse repeated decorative symbols into single characters (e.g., "———" becomes "-", "========" becomes "=").

(e) Remove residual HTML table borders and pipe characters.

(f) Replace all newline characters with spaces.

(g) Collapse multiple consecutive spaces into a single space.

(h) Trim leading and trailing whitespace from each section.

C Structured Information Extraction Questionnaire

This section presents the full questionnaire given to the language model, reformatted from the original JSON file for readability; the original questionnaire is provided as a JSON file in the data package. Each disclosure is processed with the 50 questions below, and each question corresponds to one variable in the panel. Every question follows the same structure. The first line gives, in bold, the *prompt* presented to the model, with the corresponding *variable name* set in typewriter font at the right margin. The lines beneath report the response specification—**Type**, **Format**, **Units**, and whether a response is **Required**—followed, where relevant, by the allowed range (**Validation**) and the permitted categories (**Categories**). The last line, **AI Instruction**, gives the exact instructions the model was told to follow for that question.

The variables are grouped into thematic panels by topic. Where a single topic gives rise to more than one distinct construct, each is recorded as a separate variable and placed under its respective panel; for example, capital expenditure appears both as a quantitative forecast (`capex_guidance_millions`, under Quantitative Guidance) and as a qualitative strategic descriptor (`capex_strategic_focus`, under Investment and Innovation). Similar splits apply to inventory, labor, and costs.

The prompt and the variable name serve distinct roles and are placed on opposite sides of the first line to keep them visually separate. The prompt (bold, left) is the literal question the model answers; it defines the construct being measured and is shown exactly as the model receives it. The variable name (typewriter, right) is the identifier under which the extracted value is stored in the panel and is the label used in the tables and regressions.

For every question, the model returns four outputs. `answer` is the extracted value. `status` records how the value was obtained: stated explicitly in the disclosure (EXPLICIT), inferred from context (IMPLIED), not addressed in the disclosure (NOT_MENTIONED), or not applicable to the firm (NOT_APPLICABLE). `evidence` is the supporting text drawn from the disclosure, and `confidence` is the model's self-assessed reliability (HIGH, MEDIUM, or LOW). Together, `status` and `confidence` let inferred or unsupported entries be identified. These four fields are illustrated under Q1 and omitted from the remaining questions for brevity; their full definitions and the rules governing them are given in the prompt design in Section 3.4.

Panel A. Document Identifiers

Q1. Company legal name (exact as in filing) `company_name`

Type: Text **Format:** Text **Required:** Yes
Validation: Max length 100

AI Instruction: Use exact company name from document header or first page. No abbreviations or common names. If not explicitly stated, use financial analyst knowledge to infer the legal name from context. Mark confidence as LOW and status as IMPLIED for inferences.

Recorded outputs (every item): `answer` (extracted value), `status` (EXPLICIT/IMPLIED/NOT_MENTIONED /NOT_APPLICABLE), `evidence` (supporting text), `confidence` (HIGH/MEDIUM/LOW).

Q2. Document type `report_type`

Type: Categorical (single select) **Required:** Yes
Categories: 10-Q, 10-K, 8-K, EARNINGS_RELEASE, TRANSCRIPT, OTHER

AI Instruction: Select the exact document type. Use OTHER only if none of the standard options apply. If not explicitly stated, use financial analyst knowledge to infer document type from context and content. Mark confidence as LOW and status as IMPLIED for inferences.

Q3. Official filing/release date `filing_date`

Type: Date **Format:** Date (YYYY-MM-DD) **Required:** Yes

AI Instruction: Use the official SEC filing date or press release date, not period end date. If not explicitly stated, use financial analyst knowledge to infer a reasonable date based on document metadata and context. Mark confidence as LOW and status as IMPLIED for inferences.

Q4. Financial period covered `reporting_period`

Type: Categorical (single select) **Required:** Yes
Categories: Q1, Q2, Q3, Q4, FY, OTHER

AI Instruction: Q1-Q4 for quarterly reports, FY for annual reports. If not explicitly stated, use financial analyst knowledge to infer the reporting period from context and financial statement periods. Mark confidence as LOW and status as IMPLIED for inferences.

Q5. Fiscal year `fiscal_year`

Type: Integer (numeric) **Format:** Year (YYYY) **Required:** Yes
Validation: Min 2000, Max 2030

AI Instruction: Use fiscal year, not calendar year if different. If not explicitly stated, use financial analyst knowledge to infer fiscal year from document context and reporting periods. Mark confidence as LOW and status as IMPLIED for inferences.

Panel B. Quantitative Guidance

Q6. Revenue growth guidance for next period (%) `revenue_growth_guidance_pct`

Type: Decimal (numeric) **Format:** Percentage, 1 decimal place **Units:** Percentage **Required:** Yes
Validation: Min −50, Max 100

AI Instruction: Only explicit guidance or clear directional statements. Format as X.X (e.g., 5.2 for 5.2%). If not explicitly stated, use financial analyst knowledge to infer reasonable guidance based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q7. Operating/gross margin guidance (%) `margin_guidance_pct`

Type: Decimal (numeric) **Format:** Percentage, 1 decimal place **Units:** Percentage **Required:** Yes
Validation: Min −100, Max 100

AI Instruction: Prefer operating margin, then gross margin. Format as X.X%. If not explicitly stated, use financial analyst knowledge to infer reasonable margin guidance based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q8. Capital expenditure guidance (millions USD) `capex_guidance_millions`

Type: Decimal (numeric) **Format:** Millions of USD, 1 decimal place **Units:** Millions USD **Required:** Yes
Validation: Min 0

AI Instruction: Convert to millions if needed (e.g., $1.5B = 1500.0). If not explicitly stated, use financial analyst knowledge to infer reasonable capex guidance based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q9. Expected input cost inflation (%) `input_cost_inflation_pct`

Type: Decimal (numeric) **Format:** Percentage, 1 decimal place **Units:** Percentage **Required:** Yes
Validation: Min −50, Max 50

AI Instruction: Material, labor, or energy cost inflation expectations. Format as X.X%. If not explicitly stated, use financial analyst knowledge to infer reasonable inflation expectations based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q10. Expected wage/labor cost growth (%) `wage_growth_expectation_pct`

Type: Decimal (numeric) **Format:** Percentage, 1 decimal place **Units:** Percentage **Required:** Yes
Validation: Min −30, Max 30

AI Instruction: Explicit wage growth expectations or labor cost pressures. Format as X.X%. If not explicitly stated, use financial analyst knowledge to infer reasonable wage growth expectations based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Panel C. Operating Plans and Business Conditions

Q11. Planned inventory adjustment `inventory_plan`

Type: Categorical (single select) **Required:** Yes
Categories: INCREASE, MAINTAIN, DECREASE

AI Instruction: Based on explicit inventory strategy statements. If not explicitly stated, use financial analyst knowledge to infer reasonable inventory plans based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q12. Product/service pricing direction `pricing_direction`

Type: Categorical (single select) **Required:** Yes
Categories: INCREASE, MAINTAIN, DECREASE

AI Instruction: Intended price changes or pricing power comments. If not explicitly stated, use financial analyst knowledge to infer reasonable pricing direction based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q13. Production/output change plan `production_output_plan`

Type: Categorical (single select) **Required:** Yes
Categories: INCREASE, MAINTAIN, DECREASE

AI Instruction: Planned changes in production capacity or output levels. If not explicitly stated, use financial analyst knowledge to infer reasonable production plans based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q14. Hiring/employment intentions `hiring_intentions`

Type: Categorical (single select) **Required:** Yes
Categories: EXPANDING, MAINTAINING, REDUCING

AI Instruction: Stated hiring plans or workforce changes. If not explicitly stated, use financial analyst knowledge to infer reasonable hiring intentions based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q15. Operating capacity constraints mentioned `capacity_constraints_flag`

Type: Categorical (single select) **Required:** Yes
Categories: YES, NO

AI Instruction: YES if capacity limitations/constraints mentioned, NO if explicitly stated as having adequate capacity. If not explicitly discussed, use financial analyst knowledge to infer capacity constraint status based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q16. Customer demand outlook assessment `customer_demand_outlook`

Type: Categorical (single select) **Required:** Yes

Categories: STRONG, MODERATE, WEAK, UNCERTAIN

AI Instruction: Based on management's characterization of demand trends. If not explicitly discussed, use financial analyst knowledge to infer reasonable demand outlook based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q17. Market share trend indication `market_share_direction`

Type: Categorical (single select) **Required:** Yes

Categories: GAINING, STABLE, LOSING

AI Instruction: Explicit or implied market share trends. If not explicitly mentioned, use financial analyst knowledge to infer reasonable market share direction based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q18. Competitive environment assessment `competitive_intensity`

Type: Categorical (single select) **Required:** Yes

Categories: INTENSIFYING, STABLE, MODERATING

AI Instruction: Management's view on competitive pressures. If not explicitly discussed, use financial analyst knowledge to infer reasonable competitive intensity assessment based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q19. Pricing power trend `pricing_power_trend`

Type: Categorical (single select) **Required:** Yes

Categories: STRENGTHENING, STABLE, WEAKENING

AI Instruction: Ability to maintain or raise prices. If not explicitly mentioned, use financial analyst knowledge to infer reasonable pricing power trends based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Panel D. Supply Chain

Q20. Supply chain normalization status `supply_chain_status`

Type: Categorical (single select) **Required:** Yes

Categories: NORMALIZED, IMPROVING, STILL_DISRUPTED, WORSENING

AI Instruction: Overall supply chain condition assessment. If not explicitly discussed, use financial analyst knowledge to infer reasonable supply chain status based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q21. Supplier delivery time trends `supplier_delivery_times`

Type: Categorical (single select) **Required:** Yes
Categories: IMPROVING, STABLE, WORSENING

AI Instruction: Lead time or delivery performance trends. If not explicitly mentioned, use financial analyst knowledge to infer reasonable supplier delivery trends based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q22. Freight/logistics cost trends `freight_logistics_costs`

Type: Categorical (single select) **Required:** Yes
Categories: DECREASING, STABLE, INCREASING

AI Instruction: Transportation and logistics cost trends. If not explicitly mentioned, use financial analyst knowledge to infer reasonable freight/logistics cost trends based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q23. Inventory levels relative to demand `inventory_levels_vs_demand`

Type: Categorical (single select) **Required:** Yes
Categories: EXCESS, BALANCED, INSUFFICIENT

AI Instruction: Management's characterization of inventory adequacy. If not explicitly discussed, use financial analyst knowledge to infer reasonable inventory levels relative to demand based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q24. Supply chain risk concern level `supply_chain_risk_level`

Type: Integer (numeric) **Format:** Integer scale, 1–5 **Required:** Yes
Validation: Min 1, Max 5

AI Instruction: 1=No concern, 2=Low, 3=Moderate, 4=High, 5=Severe concern. If not explicitly addressed, use financial analyst knowledge to infer reasonable supply chain risk level based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Panel E. Risk Exposures

Q25. Regulatory/compliance risk level `regulatory_risk_level`

Type: Integer (numeric) **Format:** Integer scale, 1–5 **Required:** Yes
Validation: Min 1, Max 5

AI Instruction: 1=Low risk, 5=High risk. Consider regulatory changes, compliance costs. If not explicitly mentioned, use financial analyst knowledge to infer reasonable regulatory risk level based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q26. Labor shortage/availability risk level `labor_shortage_risk_level`

Type: Integer (numeric) **Format:** Integer scale, 1–5 **Required:** Yes
Validation: Min 1, Max 5

AI Instruction: 1=No shortage concerns, 5=Severe shortage concerns. If not explicitly mentioned, use financial analyst knowledge to infer reasonable labor shortage risk level based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q27. Interest rate sensitivity level `interest_rate_sensitivity_level`

Type: Integer (numeric) **Format:** Integer scale, 1–5 **Required:** Yes
Validation: Min 1, Max 5

AI Instruction: 1=Not sensitive, 5=Highly sensitive to rate changes. If not explicitly discussed, use financial analyst knowledge to infer reasonable interest rate sensitivity based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q28. Foreign exchange risk exposure `foreign_exchange_exposure`

Type: Categorical (single select) **Required:** Yes
Categories: MINIMAL, MODERATE, SIGNIFICANT

AI Instruction: Based on international revenue/cost exposure. If not explicitly mentioned, use financial analyst knowledge to infer reasonable foreign exchange exposure based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Panel F. Macroeconomic and Policy Exposure

Q29. Monetary policy/interest rate mentions `monetary_policy_mentions`

Type: Categorical (single select) **Required:** Yes
Categories: NONE, FEW, SEVERAL, FREQUENT

AI Instruction: NONE=0, FEW=1-2, SEVERAL=3-5, FREQUENT=6+ mentions. If not explicitly clear, use financial analyst knowledge to infer reasonable monetary policy mention frequency based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q30. Tone toward monetary policy impact `monetary_policy_tone`

Type: Categorical (single select) **Required:** Yes
Categories: POSITIVE, NEUTRAL, NEGATIVE

AI Instruction: How monetary policy is viewed as affecting the business. If not explicitly discussed, use financial analyst knowledge to infer reasonable monetary policy tone based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q31. Fiscal policy/tax/subsidy mentions `fiscal_policy_mentions`

Type: Categorical (single select) **Required:** Yes
Categories: NONE, FEW, SEVERAL, FREQUENT

AI Instruction: References to government spending, taxes, subsidies. NONE=0, FEW=1-2, SEVERAL=3-5, FREQUENT=6+. If not explicitly clear, use financial analyst knowledge to infer reasonable fiscal policy mention frequency based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q32. Trade policy/tariff concern level `trade_policy_concern_level`

Type: Integer (numeric) **Format:** Integer scale, 1–5 **Required:** Yes
Validation: Min 1, Max 5

AI Instruction: 1=No concern, 5=High concern about trade policies/tariffs. If not explicitly mentioned, use financial analyst knowledge to infer reasonable trade policy concern level based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Panel G. Investment and Innovation

Q33. R&D investment direction `rd_investment_direction`

Type: Categorical (single select) **Required:** Yes
Categories: INCREASING, STABLE, DECREASING

AI Instruction: Based on R&D spending plans or emphasis. If not explicitly discussed, use financial analyst knowledge to infer reasonable R&D investment direction based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q34. Innovation/technology emphasis level `innovation_emphasis_level`

Type: Integer (numeric) **Format:** Integer scale, 1–5 **Required:** Yes
Validation: Min 1, Max 5

AI Instruction: 1=Minimal mention, 5=Heavy emphasis on innovation. Count innovation, R&D, new product mentions. If not explicitly clear, use financial analyst knowledge to infer reasonable innovation emphasis level based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q35. Digital transformation/AI mentions `digital_transformation_mentions`

Type: Categorical (single select) **Required:** Yes
Categories: NONE, FEW, SEVERAL, FREQUENT

AI Instruction: References to AI, automation, digitization. NONE=0, FEW=1-2, SEVERAL=3-5, FREQUENT=6+. If not explicitly clear, use financial analyst knowledge to infer reasonable digital transformation mention frequency based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q36. Primary CapEx strategic focus `capex_strategic_focus`

Type: Categorical (single select) **Required:** Yes

Categories: CAPACITY_EXPANSION, TECHNOLOGY_UPGRADE, EFFICIENCY, MAINTENANCE, MIXED

AI Instruction: Primary purpose of capital investments if discussed. If not explicitly specified, use financial analyst knowledge to infer reasonable capex strategic focus based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Panel H. Financial Position and Capital Allocation

Q37. Management's liquidity position characterization `liquidity_position_characterization`

Type: Categorical (single select) **Required:** Yes

Categories: STRONG, ADEQUATE, TIGHT, CONCERNING

AI Instruction: How management describes cash position, credit facilities. If not explicitly discussed, use financial analyst knowledge to infer reasonable liquidity position characterization based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q38. Management concern about debt levels `debt_level_concern`

Type: Categorical (single select) **Required:** Yes

Categories: NO_CONCERN, MONITORING, MODERATE_CONCERN, HIGH_CONCERN

AI Instruction: Tone regarding debt burden or leverage. If not explicitly addressed, use financial analyst knowledge to infer reasonable debt level concern based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q39. Shareholder return priority emphasis `shareholder_return_priority`

Type: Categorical (single select) **Required:** Yes

Categories: DIVIDENDS, BUYBACKS, BOTH, GROWTH_INVESTMENT

AI Instruction: Capital allocation priority if discussed. If not explicitly mentioned, use financial analyst knowledge to infer reasonable shareholder return priority based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Panel I. Management Sentiment and Tone

Q40. Overall management sentiment score `overall_management_sentiment`

Type: Integer (numeric) **Format:** Integer scale, 1–5 **Required:** Yes

Validation: Min 1, Max 5

AI Instruction: 1=Very negative, 2=Negative, 3=Neutral, 4=Positive, 5=Very positive. If not explicitly clear, use financial analyst knowledge to infer reasonable management sentiment based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q41. Management confidence in achieving guidance `confidence_in_guidance`

Type: Integer (numeric) **Format:** Integer scale, 1–5 **Required:** Yes
Validation: Min 1, Max 5

AI Instruction: 1=Low confidence, 5=High confidence. If not explicitly clear, use financial analyst knowledge to infer reasonable confidence in guidance based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q42. Uncertainty level expressed `uncertainty_level_expressed`

Type: Integer (numeric) **Format:** Integer scale, 1–5 **Required:** Yes
Validation: Min 1, Max 5

AI Instruction: 1=Very certain, 5=Highly uncertain. Count hedging language, 'challenging to predict', etc. If not explicitly clear, use financial analyst knowledge to infer reasonable uncertainty level based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q43. Forward visibility characterization `forward_visibility_characterization`

Type: Categorical (single select) **Required:** Yes
Categories: CLEAR, REASONABLE, LIMITED, POOR

AI Instruction: Management's characterization of ability to predict future performance. If not explicitly discussed, use financial analyst knowledge to infer reasonable forward visibility characterization based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Panel J. Narrative Variables

Q44. Key positive business drivers mentioned `key_positive_drivers`

Type: Free text **Required:** Yes

AI Instruction: Summarize main positive factors mentioned by management. Maximum 300 characters. If not explicitly stated, use financial analyst knowledge to infer reasonable positive drivers based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q45. Key risk factors or headwinds mentioned `key_risk_factors`

Type: Free text **Required:** Yes

AI Instruction: Summarize main risks or challenges mentioned. Maximum 300 characters. If not explicitly stated, use financial analyst knowledge to infer reasonable risk factors based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q46. Most representative management quote `representative_quote`

Type: Free text **Required:** Yes

AI Instruction: Direct quote that best captures management sentiment. Maximum 200 characters including quotes. If no suitable quote available, use financial analyst knowledge to create a representative summary based on company context and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q47. One-sentence analyst interpretation `analyst_interpretation_summary`

Type: Free text **Required:** Yes

AI Instruction: Concise summary of company outlook and position. Maximum 200 characters. If not explicitly clear, use financial analyst knowledge to infer reasonable interpretation based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Panel K. Geographic and Sector Classification

Q48. Geopolitical risk exposure regions `geopolitical_risk_regions`

Type: Categorical (multi-select) **Required:** Yes

Categories: CHINA, EUROPE, RUSSIA_UKRAINE, MIDDLE_EAST, ASIA_PACIFIC, LATIN_AMERICA, OTHER

AI Instruction: Select all regions mentioned as sources of geopolitical risk or uncertainty. If not explicitly mentioned, use financial analyst knowledge to infer reasonable geopolitical risk regions based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q49. Primary geographic revenue exposure `primary_geographic_exposure`

Type: Categorical (multi-select) **Required:** Yes

Categories: US, CANADA, MEXICO, EUROPE, UK, CHINA, JAPAN, ASIA_PACIFIC, MIDDLE_EAST, AFRICA, LATIN_AMERICA

AI Instruction: Select all regions with significant revenue exposure mentioned. If not explicitly stated, use financial analyst knowledge to infer reasonable geographic exposure based on company context, industry norms, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

Q50. Primary business sector/industry `sector_or_industry`

Type: Categorical (single select) **Required:** Yes

Categories: TECHNOLOGY, HEALTHCARE, FINANCIALS, INDUSTRIALS, CONSUMER_DISCRETIONARY, CONSUMER_STAPLES, ENERGY, MATERIALS, UTILITIES, REAL_ESTATE, TELECOMMUNICATIONS, OTHER

AI Instruction: Select primary sector based on main business activities. If not explicitly clear, use financial analyst knowledge to infer reasonable sector/industry classification based on company context, business description, and available information. Mark confidence as LOW and status as IMPLIED for inferences.

D Full List of Extracted Variables

Table D.1: List of extracted variables by group

#	Variable	Description	Type
Panel A: Document Identifiers			
1	`company_name`	Company legal name (exact as in filing)	Text
2	`report_type`	Document type	Categorical
3	`filing_date`	Official filing/release date	Date
4	`reporting_period`	Financial period covered	Categorical
5	`fiscal_year`	Fiscal year	Year
Panel B: Quantitative Guidance			
6	`revenue_growth_guidance_pct`	Revenue growth guidance for next period (%)	Decimal (%)
7	`margin_guidance_pct`	Operating/gross margin guidance (%)	Decimal (%)
8	`capex_guidance_millions`	Capital expenditure guidance (millions USD)	Decimal (USD mn)
9	`input_cost_inflation_pct`	Expected input cost inflation (%)	Decimal (%)
10	`wage_growth_expectation_pct`	Expected wage/labor cost growth (%)	Decimal (%)
Panel C: Operating Plans & Business Conditions			
11	`inventory_plan`	Planned inventory adjustment	Categorical
12	`pricing_direction`	Product/service pricing direction	Categorical
13	`production_output_plan`	Production/output change plan	Categorical
14	`hiring_intentions`	Hiring/employment intentions	Categorical
15	`capacity_constraints_flag`	Operating capacity constraints mentioned	Categorical
16	`customer_demand_outlook`	Customer demand outlook assessment	Categorical
17	`market_share_direction`	Market share trend indication	Categorical
18	`competitive_intensity`	Competitive environment assessment	Categorical
19	`pricing_power_trend`	Pricing power trend	Categorical
Panel D: Supply Chain			
20	`supply_chain_status`	Supply chain normalization status	Categorical
21	`supplier_delivery_times`	Supplier delivery time trends	Categorical
22	`freight_logistics_costs`	Freight/logistics cost trends	Categorical
23	`inventory_levels_vs_demand`	Inventory levels relative to demand	Categorical
24	`supply_chain_risk_level`	Supply chain risk concern level	Ordinal 1–5
Panel E: Risk Exposures			
25	`regulatory_risk_level`	Regulatory/compliance risk level	Ordinal 1–5
26	`labor_shortage_risk_level`	Labor shortage/availability risk level	Ordinal 1–5

continued on next page

Table D.1 (continued)

#	Variable	Description	Type
27	`interest_rate_sensitivity_level`	Interest rate sensitivity level	Ordinal 1–5
28	`foreign_exchange_exposure`	Foreign exchange risk exposure	Categorical
Panel F: Macroeconomic & Policy Exposure			
29	`monetary_policy_mentions`	Monetary policy/interest rate mentions	Categorical
30	`monetary_policy_tone`	Tone toward monetary policy impact	Categorical
31	`fiscal_policy_mentions`	Fiscal policy/tax/subsidy mentions	Categorical
32	`trade_policy_concern_level`	Trade policy/tariff concern level	Ordinal 1–5
Panel G: Investment & Innovation			
33	`rd_investment_direction`	R&D investment direction	Categorical
34	`innovation_emphasis_level`	Innovation/technology emphasis level	Ordinal 1–5
35	`digital_transformation_mentions`	Digital transformation/AI mentions	Categorical
36	`capex_strategic_focus`	Primary CapEx strategic focus	Categorical
Panel H: Financial Position & Capital Allocation			
37	`liquidity_position_characterization`	Management's liquidity position characterization	Categorical
38	`debt_level_concern`	Management concern about debt levels	Categorical
39	`shareholder_return_priority`	Shareholder return priority emphasis	Categorical
Panel I: Management Sentiment & Tone			
40	`overall_management_sentiment`	Overall management sentiment score	Ordinal 1–5
41	`confidence_in_guidance`	Management confidence in achieving guidance	Ordinal 1–5
42	`uncertainty_level_expressed`	Uncertainty level expressed	Ordinal 1–5
43	`forward_visibility_characterization`	Forward visibility characterization	Categorical
Panel J: Narrative Variables			
44	`key_positive_drivers`	Key positive business drivers mentioned	Free text
45	`key_risk_factors`	Key risk factors or headwinds mentioned	Free text
46	`representative_quote`	Most representative management quote	Free text
47	`analyst_interpretation_summary`	One-sentence analyst interpretation	Free text
Panel K: Geographic & Sector Classification			
48	`geopolitical_risk_regions`	Geopolitical risk exposure regions	Categorical
49	`primary_geographic_exposure`	Primary geographic revenue exposure	Categorical
50	`sector_or_industry`	Primary business sector/industry	Categorical

Notes: The variables are grouped into thematic panels by topic. Where a topic admits both a quantitative forecast and a qualitative descriptor, each is listed separately under its respective group—for instance, capital expenditure appears as `capex_guidance_millions` (Panel B) and `capex_strategic_focus` (Panel G). The # column gives the variable ID, which corresponds to the question number in Appendix C.

E Foundation for Variable Selection

This section sets out the justification for each extracted variable along two dimensions: its basis in SEC disclosure requirements and its support in the academic literature. I impose both requirements on every variable. Each must correspond to information that registrants are required to disclose if applicable, so that the variable is reliably present across filings; and each must have established relevance in prior research on the information content of corporate disclosure, so that it is economically meaningful. Appendix E.1 documents, for every variable, the governing provision of Regulation S-K (with the operative regulatory text and a rationale), and Appendix E.2 documents the supporting literature.

E.1 Evidence from SEC Regulation S-K

Table E.1 outlines the statutory and regulatory basis for our variable selection, drawing evidence directly from U.S. Securities and Exchange Commission (SEC) Regulation S-K (17 CFR Part 229) (U.S. Securities and Exchange Commission, 2023). It systematically maps each extracted variable to the specific SEC mandates and principles-based reporting requirements that compel management to discuss these economic themes. By linking our dataset directly to the regulatory architecture of 10-K and 10-Q filings, this table validates that our large language model targets high-density, standardized financial narratives rather than isolated or ad-hoc corporate commentary.

Table E.1: Regulatory anchors for variable extraction

SEC Item	**Variables (IDs)**	**Supporting eCFR Quote and Macro Rationale**
Item 101 (Business)	10, 14, 26	*"A description of the registrant's human capital resources [. . .] including any human capital measures or objectives that the registrant focuses on [. . .] attraction and retention of personnel."* (17 CFR § 229.101(c)(2)(ii)). **Rationale:** This mandate requires firms to discuss their ability to find and keep workers. In a macro context, mentions of "attraction and retention" challenges serve as a proxy for labor market tightness and wage-push inflation pressures.
Item 101 (Business)	33, 34	*"Status of development efforts for new or enhanced products [. . .]"* (17 CFR § 229.101(c)(1)(ii)). **Rationale:** Grounds innovation and R&D investment in reported financial commitments to track productivity trends.
Item 101	8, 16, 17, 18, 25, 36, 50	*"market demand and competitive conditions"* (17 CFR § 229.101(c)(1)(ii)).

Continued on next page

Table E.1 – continued from previous page

SEC Item	Variables (IDs)	Supporting eCFR Quote and Macro Rationale
		"The material effects that compliance with Government profits [. . .] may have upon capital expenditures, earnings and competitive position." (17 CFR § 229.101(c)(2)(i)).
(Business)		**Rationale:** It explicitly mentions trends in market demand, competitive conditions, and material effects of government regulation.
Item 303	6, 7, 13, 15	*"Describe any known trends or uncertainties that have had or that are reasonably likely to have a material [. . .] impact on net sales or revenues or income from continuing operations [. . .]"* (17 CFR § 229.303(b)(2)(ii)).
(MD&A)		**Rationale:** The legal basis for revenue and margin guidance, operation and production plan.
Item 303	9, 10, 11, 12, 19, 20, 21, 22, 23, 24	*"If the registrant knows [. . .] a material change in the relationship between costs and revenues (such as [. . .] future increases in costs of labor or materials or price increases or inventory adjustments"* (17 CFR § 229.303(b)(2)(ii)).
(MD&A)		**Rationale:** This is the fundamental basis for the pricing power, supply conditions, and input inflation related variables. Management is legally compelled to disclose if they expect costs to rise and whether they intend to adjust prices to compensate, providing a direct link to the New Keynesian Phillips Curve. It also supports tracking the inventory cycle, a key component of quarterly GDP volatility. Inventory-demand mismatches are standard narrative explanations for performance and outlook.
Item 303	8, 36, 37, 38, 39	*"Describe the registrant's material cash requirements, including commitments for capital expenditures [. . .] the anticipated source of funds [. . .] and the general purpose of such commitments."* (17 CFR § 229.303(b)(1)(ii)). *"Describe any known material trends [. . .] in the [. . .] capital resources. Indicate [. . .] changes in the mix and relative cost of such resources [. . .]"* (17 CFR § 229.303(b)(1)(ii)).
(MD&A)		**Rationale:** This is the core SEC source for liquidity, capital resources, funding needs, and the purpose of major cash requirements.

Continued on next page

Table E.1 – continued from previous page

SEC Item	Variables (IDs)	Supporting eCFR Quote and Macro Rationale
Item 303 (MD&A)	16, 37	*"Identify any known trends or any known demands [. . .] result in the registrant's liquidity increasing or decreasing in any material way [. . .] Also identify and separately describe internal and external sources of liquidity, and briefly discuss any material unused sources of liquid assets."* (17 CFR § 229.303(b)(1)(i)). **Rationale:** It directly requires discussion of liquidity, sources of liquidity, and known trends affecting liquidity.
Item 303 (MD&A)	38, 40, 41, 42, 43	*"The objective [. . .] is to provide [. . .] a narrative explanation of the financial statements [. . .] that allows investors to view the registrant from the perspective of management."* (17 CFR § 229.303(a)). **Rationale:** Justifies the use of LLM to extract sentiment and uncertainty directly from the management's unique perspective.
Item 105 (Risk Factors)	24, 27, 29, 30, 31, 32	*"Provide [. . .] a discussion of the material factors that make an investment in the registrant speculative or risky. [. . .] explain how each risk affects the registrant."* (17 CFR § 229.105(a)(b)). **Rationale:** Anchors all macro-thematic risks (Monetary Policy, Trade, Fiscal Policy). The mandate to explain "how" the risk affects them forces non-generic disclosure.
Item 305 (Market Risk)	27, 28	*"Provide [. . .] quantitative information about market risk [. . .] i.e., interest rate risk, foreign currency exchange rate risk [. . .]"* (17 CFR § 229.305(a)(1)). **Rationale:** Direct source for Interest Rate and FX sensitivity scores based on mandatory quantitative modeling by the firm.
Item 703 (Issuer Repurchases)	38	*"Each plan or program the issuer has determined to terminate prior to expiration, or under which the issuer does not intend to make further purchases"* (7 CFR § 229.303(b)(4)). **Rationale:** It requires issuers to disclose the nature of their repurchase plans and any intent to cease future purchases.

Continued on next page

Table E.1 – continued from previous page

SEC Item	Variables (IDs)	Supporting eCFR Quote and Macro Rationale
Others	35	*"if material, report separately on how issuers are deploying AI and the effects of AI deployment on (a) internal business operations and (b) consumer facing matters"* (U.S. Securities and Exchange Commission (2024)). **Rationale:** AI-related 10-K disclosures have surged, with firms expected to substantiate AI claims and disclose material risks.

E.2 Academic Literature and Macroeconomic Rationale

Table E.2 details the academic literature supporting the economic validity and predictive power of the extracted variables. Anchored by the Granular Hypothesis (Gabaix, 2011), which posits that macroeconomic fluctuations are significantly driven by idiosyncratic firm-level shocks, this table demonstrates how our micro-level variables could serve as leading indicators for broader macro aggregates. The justification draws heavily on recent empirical advancements in text analysis and natural language processing within finance. Specifically, I highlight empirical studies that successfully extract highly similar economic variables—such as management sentiment—from corporate disclosures like earnings conference calls. I extend this approach to the mandatory 10-K and 10-Q disclosures, selecting variables that prior work has shown to be both extractable from corporate text and economically informative.

Table E.2: Paper-based rationale for disclosure variables

Paper source	Variable IDs	Why this paper rationalizes the variables
Davis et al. (2012), Jha et al. (2024), Bonsall IV et al. (2013), Gennaioli et al. (2016)	6, 7, 8	Validates that both qualitative corporate text and managerial forecasts inherently capture undiversifiable macroeconomic risks. Applying text analysis to extract forward-looking expectations on revenue, margins, and capital expenditures generates a robust, real-time leading indicator for imminent shifts in aggregate output and private domestic investment.
Windsor and Zang (2023)	9, 10, 12, 16, 19, 22	This paper constructs firm-level text indices for input costs, demand, and final prices from earnings calls. It directly motivates extracting expected input-cost inflation, product/service pricing direction, customer-demand outlook, pricing-power trend, and freight/logistics cost pressure from corporate narratives.
Blinder et al. (1981), Khan and Thomas (2007)	22, 23	Blinder and Maccini (1991) establish that coordinated firm-level inventory shifts are the primary "swing factor" driving aggregate macroeconomic business cycles. Within this macro framework, Khan and Thomas (2007) demonstrate that fixed ordering and delivery costs dictate these inventory cycles via an (S, s) policy, where firms delay ordering until inventory falls to a specific lower threshold (s) before triggering a massive restocking order (S)

Paper source	Variable IDs	Why this paper rationalizes the variables
Acemoglu et al. (2012)	13, 20, 21, 24	It demonstrates mathematically that idiosyncratic, firm-level shocks do not cancel out, but instead propagate through asymmetric input-output networks to generate aggregate macroeconomic fluctuations. Within this theoretical framework, production change plans capture the origin of localized, node-level shocks, while managerial assessments of supply chain normalization, delivery times, and supply chain risks measure the structural network frictions and transmission mechanisms that govern how broadly these micro-shocks cascade into systemic economic cycles.
Gosselin and Taskin (2023), Ulrich Ruch and Taskin (2024)	15, 16, 20, 21, 24	Demonstrates that NLP-extracted mentions of customer demand, capacity constraints, and supply bottlenecks capture real-time macroeconomic supply and demand shocks. These variables serve as high-frequency indicators for the macroeconomic output gap and structural inflation.
Albrizio et al. (2023)	9, 29, 30	This paper extracts firms' inflation expectations from earnings calls and links them to future inflation and policy transmission. It supports expected input-cost inflation as a forward-looking expectation variable and also supports monetary-policy/interest-rate mentions and tone toward monetary-policy impact when those discussions are embedded in the disclosure.
Kalyani and Ozkan (2025)	10, 14, 26	This paper measures firm-level labor-cost pressures and inflation pass-through from text. It provides strong support for expected wage/labor-cost growth, hiring/employment intentions, and labor-shortage or labor-availability risk as text-based corporate constructs.
Cortes et al. (2023)	29, 30	This paper shows that macroeconomic issues are frequently discussed in conference calls and convey monetary-policy information. It is a direct rationale for coding monetary-policy/interest-rate mentions and the tone of management's discussion of monetary-policy effects.
Hassan et al. (2019)	25, 29, 30, 31, 32	This paper constructs a firm-level political-risk measure from earnings calls. It supports coding policy-sensitive discussion as structured variables, including regulatory/compliance risk, fiscal-policy/tax/subsidy mentions, and trade-policy/tariff concern.

Paper source	Variable IDs	Why this paper rationalizes the variables
Hoberg and Phillips (2016)	12, 17, 18, 19, 33, 50	This paper shows that 10-K text can recover product-market similarity and industry structure. Text-based product market classifications show that firms with high pricing power and favorable competitive dynamics dictate industry margin cycles and drive macroeconomic inflation persistence.
Merkley (2014)	33, 34	This paper shows that narrative R&D disclosure is informative and economically meaningful. It supports coding R&D investment direction and innovation/technology emphasis from firms' discussion of research and development activities.
Soto (2025)	35	This paper introduces a firm-level AI R&D measure from conference-call text. It directly supports a digital-transformation/AI-mentions variable, especially if the variable is framed as AI emphasis or AI-related strategic discussion rather than hard-realized adoption.
Loughran and McDonald (2011)	30, 40, 42, 43	This paper is the foundational justification for using finance-specific text dictionaries instead of generic sentiment lexicons. It supports tone toward monetary policy impact, overall management sentiment, uncertainty level, and forward visibility as finance-specific language constructs.
Jiang et al. (2019), Druz et al. (2020)	40, 42, 43	They construct a manager-sentiment index from corporate financial disclosures and show that changes in managers' tone predict lower future earnings and higher uncertainty. It supports coding overall sentiment, expresses uncertainty, and provides forward visibility from conference-call language.
Bernanke et al. (1999), Buehlmaier and Whited (2018), Hoberg and Maksimovic (2015)	37, 38, 39	Text-based measures of liquidity and debt concerns reflect underlying financial constraints. These micro-level constraints aggregate into the "Financial Accelerator," amplifying macroeconomic recessions.
Bloom (2009)	40, 41, 42, 43	Links "soft" management sentiment and uncertainty levels to the real-options effect, predicting recessionary "wait-and-see" behavior.
Bloom et al. (2021)	33, 34, 35, 36	Uses text analysis of earnings calls to prove that corporate mentions of AI and digital transformation map directly to the creation of intangible capital and long-run shifts in Total Factor Productivity (TFP).

Paper source	Variable IDs	Why this paper rationalizes the variables
Baker et al. (2016), Caldara and Iacoviello (2022)	25, 27, 28, 29, 30, 31, 32	Proves that news and corporate text regarding monetary, fiscal, and trade policy form a measurable "uncertainty index." This uncertainty triggers a real-options effect, causing aggregate investment and hiring contractions.

Notes: Variables may appear in both the SEC-based and paper-based tables. This is intentional. The SEC table gives the disclosure-regime justification; the paper table gives the text-as-data justification for why the construct is empirically measurable and economically meaningful.

F Sample Construction

Table F.1 summarizes the filing sample and the date restrictions underlying variable selection and the pre-pandemic analysis. The full corpus also supports the subsequent evaluation through 2026.

Table F.1: Sample construction

	Filings	Firms
10-Q and 10-K filings by S&P 500 members, January 1996–September 2026	60,236	1,159
Filings without a valid structured output	0	
Filings with valid structured outputs (3,011,800 answers)	60,236	1,159
Filed before 2020 (pre-pandemic analysis)	46,485	1,029
Filed before 2013 (variable selection)	32,475	877

Notes: A filing enters the sample if the firm was an S&P 500 member on the filing date. Each filing is dated by its EDGAR filing date and counted once; a filing matched to two fiscal quarters (for example, when a fiscal year ends in late January or February) is assigned to one of them.

G Supplementary Evidence for GDP Forecasting

G.1 Cross-Correlations and Lead–Lag Relationships

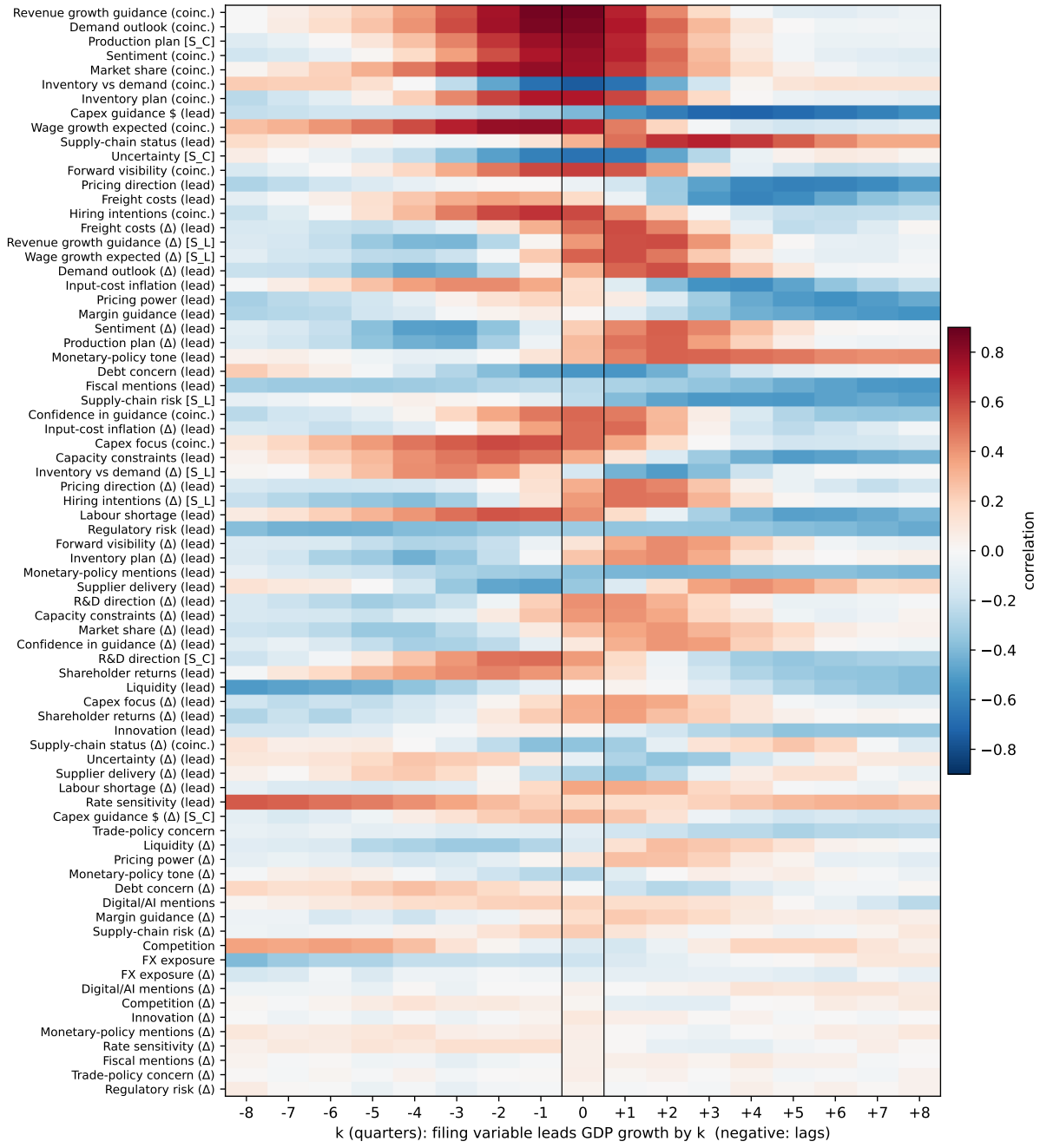


Figure G.1: Cross-Correlations of Filing-Based Measures with GDP Growth

Notes: Cross-correlations of the filing-based measures, in levels and as quarterly changes (Δ), with year-over-year real GDP growth (first release), training sample 1996Q2–2012Q4. Positive k: filing variable leads GDP; negative k: GDP leads. Rows are sorted by peak absolute correlation; [S_L] and [S_C] mark the selected leading and coincident variables, (lead) and (coinc.) the other CCF candidates.

Table G.1: Lead/lag coefficients: real GDP growth (YoY, first release) on filing-based measures in levels, training sample 1996Q2–2012Q4

Variable	9m before	6m before	3m before	Concurrent	3m after	6m after	9m after
Likert variables (0–1 scale)							
Supply-chain risk	**-27.95*****	-25.71**	-21.30**	-12.67*	-6.24	-1.47	0.64
Regulatory risk	**-12.73*****	-12.51**	-12.69**	-11.82**	-12.55**	-13.75**	-15.64**
Labour shortage	-10.51**	-2.49	6.03	15.16*	19.09**	20.65***	**18.28*****
Rate sensitivity	11.08*	8.31	8.06*	9.32*	**12.91****	16.17**	19.42**
Trade-policy concern	**-15.50****	-12.59*	-11.15	-6.57	-8.66	-9.15	-10.77
Innovation	**-9.49*****	-6.47**	-3.60	1.14	1.34	0.57	-1.56
Uncertainty	-8.28	-15.01**	-20.25***	-21.78***	**-21.60*****	-17.89***	-12.94***
Sentiment	8.13*	14.39***	19.52***	22.32***	**21.11*****	16.85***	10.76***
Confidence in guidance	3.23	13.17	21.13*	23.43**	**21.22****	14.59**	7.29
Continuous variables							
Revenue growth guidance	0.12	0.30***	0.48***	**0.59*****	0.60***	0.53***	0.40***
Margin guidance	**-0.37*****	-0.24*	-0.10	-0.01	0.03	0.01	-0.04
Capex guidance $	-81.15***	-67.92***	-54.52***	-42.40***	**-33.67*****	-27.05***	-22.42***
Input-cost inflation	**-0.93*****	-0.66**	-0.19	0.26	0.59	0.73*	0.72**
Wage growth expected	-0.07	0.73*	1.72***	2.65***	**3.03*****	2.97***	2.64***
Categorical variables (0–1 scale)							
Inventory plan	4.45	9.73**	14.59***	17.69***	**17.95*****	15.31***	10.62***
Pricing direction	**-12.40*****	-8.34**	-4.11	-1.34	-0.26	-0.26	-1.18
Production plan	3.94*	7.67***	11.14***	13.34***	**12.96*****	10.86***	7.56***
Hiring intentions	-0.12	3.49	7.11*	9.97**	10.81***	10.03***	**7.66*****
Capacity constraints	-8.86**	-3.37	3.08	9.27	12.73**	14.14**	**13.15*****
Demand outlook	3.37*	5.97***	8.33***	**9.79*****	10.06***	9.01***	7.04***
Market share	14.76*	23.36**	31.09***	37.37***	**39.49*****	36.66***	30.80***
Competition	3.86	-2.63	-8.39	-6.69	-4.57	1.48	**6.86**
Pricing power	**-9.04****	-3.38	2.19	4.33	5.18	3.56	0.78
Supply-chain status	**29.80*****	27.89***	21.27**	13.61**	5.95	1.11	-1.11
Supplier delivery	38.45*	16.66	-12.74	-36.03*	-52.19***	**-49.30*****	-35.70**
Freight costs	**-8.99****	-5.52*	-0.95	2.93	5.39	6.38	5.83*
Inventory vs demand	-10.12	-25.39***	-38.30***	-43.29***	**-38.52*****	-27.06***	-12.88*
FX exposure	-3.15	-6.56	-9.74	-13.19	-15.31	-16.91	**-17.27**
Monetary-policy mentions	**-7.29*****	-7.41**	-7.30**	-6.95**	-7.04**	-7.25**	-7.07**
Monetary-policy tone	**16.78*****	16.90**	14.56*	8.98	4.25	0.49	-1.55
Fiscal mentions	**-5.92****	-5.21**	-4.63*	-4.01	-4.33*	-4.92*	-5.53*
R&D direction	-7.13*	-1.43	5.75	12.77	16.31*	15.69**	**12.32*****
Digital/AI mentions	2.42	2.98	2.89	3.62	3.35	**3.55***	3.04
Capex focus	0.03	2.91	6.13*	9.07**	10.34**	10.87***	**10.30*****
Liquidity	-5.17	-0.46	1.68	-0.36	-4.42	-9.41	**-13.41***
Debt concern	-14.33	-20.56*	-25.03**	-25.57**	-23.60***	**-18.90*****	-13.09**
Shareholder returns	-4.66	-1.00	3.82	7.90	10.49	**11.22***	10.02
Forward visibility	4.81	12.89*	19.22***	21.41***	**20.63*****	16.44***	11.10**

Notes: $y_q = \alpha + \beta\, x_{q-k} + \varepsilon_q$ estimated on 1996Q2–2012Q4, the variable-selection sample; y_q is year-over-year growth of real GDP as first published (BEA advance estimate), x the equal-weighted cross-sectional mean of the filing-based measure in the filing quarter. Each variable in levels. HAC standard errors (Newey–West, 4 lags). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Bold denotes the most significant lag for each variable.

Table G.2: Lead/lag coefficients: real GDP growth (YoY, first release) on filing variables in quarterly changes, training sample 1996Q2–2012Q4

Variable	9m before	6m before	3m before	Concurrent	3m after	6m after	9m after
Likert variables (0–1 scale)							
Supply-chain risk (Δ)	-1.93	4.27	8.75	**17.32**	12.30	8.46	3.56
Regulatory risk (Δ)	-1.21	0.16	-0.76	1.87	-2.55	-1.80	**-3.12**
Labour shortage (Δ)	10.02	**16.50****	17.84**	18.56**	7.97	2.04	-4.93
Rate sensitivity (Δ)	-8.20	-7.63	0.87	5.48	**16.33**	14.37	13.38
Trade-policy concern (Δ)	-0.36	1.91	0.94	**2.97**	-1.22	-0.04	-0.55
Innovation (Δ)	0.36	1.74	1.63	**2.71**	0.15	-0.46	-1.22
Uncertainty (Δ)	**-18.51****	-19.89*	-15.08*	-3.90	3.63	11.62	14.06
Sentiment (Δ)	29.29**	**35.87*****	28.84**	14.67	-7.09	-23.83	-33.08**
Confidence in guidance (Δ)	**31.08****	32.17**	25.76**	7.88	-6.71	-20.26	-23.22
Continuous variables							
Revenue growth guidance (Δ)	0.61***	**0.90*****	0.88***	0.59**	0.03	-0.38	-0.63
Margin guidance (Δ)	0.35	0.41	**0.47***	0.32	0.13	-0.07	-0.20
Capex guidance $ (Δ)	-0.00	0.00	0.01	0.01*	0.01*	**0.01****	0.01**
Input-cost inflation (Δ)	0.13	0.98*	1.65***	**1.62*****	1.17***	0.53	-0.01
Wage growth expected (Δ)	1.93*	3.86***	**4.77*****	4.36***	1.88	-0.19	-1.59
Categorical variables (0–1 scale)							
Inventory plan (Δ)	**15.93****	20.50**	18.97**	11.90*	0.01	-11.02	-19.75*
Pricing direction (Δ)	19.51*	34.14**	**37.03*****	24.58**	9.72	-0.03	-8.34
Production plan (Δ)	13.66*	**19.75****	18.04**	10.64	-2.18	-11.15	-16.74**
Hiring intentions (Δ)	11.75*	18.40**	18.69**	**14.46****	4.56	-3.90	-12.04
Capacity constraints (Δ)	9.38	16.53**	19.44**	**18.64****	10.19	3.66	-3.19
Demand outlook (Δ)	13.23**	16.95***	**15.27*****	9.33**	1.28	-7.09	-12.82*
Market share (Δ)	33.40	43.55**	37.38*	**29.62****	8.87	-11.07	-25.18
Competition (Δ)	**-4.19**	-4.01	-3.79	1.00	1.33	3.74	3.49
Pricing power (Δ)	8.44*	10.62*	**11.02***	4.37	1.64	-3.08	-5.43
Supply-chain status (Δ)	11.27	-7.24	-26.70*	-31.22*	**-31.71****	-19.11**	-8.58
Supplier delivery (Δ)	-9.54	-35.43*	-48.96*	-38.14*	**-26.89****	4.61	22.58
Freight costs (Δ)	6.87	17.54***	23.02***	**19.47*****	12.14***	4.68	-2.84
Inventory vs demand (Δ)	-41.71***	**-54.47*****	-46.06**	-16.67	17.07	41.07	51.29**
FX exposure (Δ)	-10.66	-10.80	-11.43	-11.79	-10.88	**-10.01**	-3.77
Monetary-policy mentions (Δ)	-1.85	-0.87	0.40	**2.15**	1.76	1.38	2.82
Monetary-policy tone (Δ)	2.85	0.27	-5.42	**-12.98***	-11.34*	-8.54*	-4.38
Fiscal mentions (Δ)	0.86	**1.66**	1.31	1.41	-1.12	-1.33	-1.50
R&D direction (Δ)	11.19	19.64*	25.62**	**25.68****	13.55	-2.04	-12.23
Digital/AI mentions (Δ)	**1.07**	0.69	-0.13	0.93	-0.51	0.01	-0.83
Capex focus (Δ)	7.22*	**12.16****	13.39**	11.83**	5.29	1.25	-3.18
Liquidity (Δ)	35.33	40.38	19.74	-19.11	-42.92	-55.04*	**-48.10****
Debt concern (Δ)	**-20.30**	-22.62	-16.42	-1.44	7.60	15.42	19.04
Shareholder returns (Δ)	13.08	16.34	**22.37**	19.78	13.75	5.47	-4.23
Forward visibility (Δ)	**23.72****	26.46**	20.61**	7.15	-2.84	-12.90	-16.27

Notes: $y_q = \alpha + \beta\, x_{q-k} + \varepsilon_q$ estimated on 1996Q2–2012Q4, the variable-selection sample; y_q is year-over-year growth of real GDP as first published (BEA advance estimate), x the equal-weighted cross-sectional mean of the filing variable in the filing quarter. Each variable as its quarter-on-quarter change. HAC standard errors (Newey–West, 4 lags). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Bold denotes the most significant lag for each variable.

H Further Evidence on the Information in the Filings

The forecasting comparisons in the main text ask whether the filing-based measures beat an autoregressive benchmark out of sample. This appendix reports four standard in-sample tests of whether they carry information about the economy beyond what established predictors and professional forecasters already use. To avoid any data-driven selection, the tests use five pre-specified quarterly measures built from the weekly economy-wide aggregates of Section 4.1 (all members' latest answers, at most 26 weeks old): an activity composite (ACT), the mean of the standardized answers on demand outlook, production plans, hiring, revenue guidance, inventory plans, sentiment, forward visibility, confidence in guidance and capacity constraints minus expressed uncertainty; its 13-week change (ACT_Δ); a price-pressure composite (PRICE) of input costs, wage growth, pricing direction, pricing power, freight costs, capacity constraints and labor shortage minus supplier delivery and supply-chain status; and the GDP leading and coincident indices of Section 5 (LEAD, COIN). All tests use 1997–2019; q indexes calendar quarters and forecast horizons h are measured in quarters.

Professional forecasters' errors. Following Coibion and Gorodnichenko (2015), Table H.1 regresses the error of the median Survey of Professional Forecasters forecast—the first release minus the forecast—on a filing-based measure known at the survey deadline. A significant coefficient means that the forecasters did not fully use information contained in the filings. For industrial production, the change in the activity composite and the leading index predict the errors of the current-quarter forecast ($t = 3.75$ and 2.97), and the change in the activity composite also those of the next-quarter forecast ($t = 2.62$): when firms' reports improve, production comes in above the forecast. For core inflation, the same two measures predict the errors with a negative sign in both quarters (t between -2.1 and -3.5): improving reports precede core inflation below the forecast. For real GDP, the levels of the activity and price composites and the coincident index predict negative errors (t of -2.1 to -2.6), so that when firms are most upbeat, the professional forecasts of growth turn out too high. For headline inflation, none of the measures is related to the errors.

Predictive regressions with standard controls. Table H.2 adds each filing-based measure to a regression of the target one, two or four quarters ahead on the latest published value and seven standard predictors: the term spread, the BAA spread, the stock market return, the VIX, Michigan consumer sentiment, the CFNAI and the NFCI. The change in the activity composite and the leading index are significant one quarter ahead for real GDP ($t = 4.5$ and 4.0), industrial production (6.3 and 4.3), business investment (7.2 and 6.9), payrolls (4.0 and 4.5) and the unemployment rate (-5.3 and -5.1), and most remain significant two quarters ahead. For core inflation, the price composite is significant at the 10% level one quarter ahead and strongly two and four quarters ahead ($t = 3.5$ and 5.5), and the change in the activity composite and the leading index enter with a negative sign at one and two quarters (t between -2.9 and -3.6). None of the measures predicts headline inflation.

Table H.1: Do filing-based measures predict professional forecasters' errors?

Target	Quarter	ACT	ACT_Δ	PRICE	LEAD	COIN	N
Real GDP	current	-2.13**	-0.02	-2.62***	-0.92	-2.06**	92
	next	-2.16**	0.66	-2.54**	-0.73	-2.11**	91
Industrial production	current	1.10	3.75***	-1.57	2.97***	0.98	92
	next	-0.23	2.62***	-1.88*	1.09	-0.29	91
Payrolls	current	1.31	1.91*	-1.18	2.25**	1.22	65
	next	0.98	1.07	-1.80*	0.66	0.96	64
Unemployment rate	current	-0.59	-1.75*	1.90*	-1.75*	-0.49	92
	next	-1.46	-2.11**	1.59	-1.95*	-1.45	91
Core CPI	current	-0.11	-2.11**	0.04	-3.03***	-0.12	52
	next	0.36	-3.45***	0.50	-2.58***	0.40	51
Headline CPI	current	0.35	0.51	-0.01	-0.33	-0.02	92
	next	-0.15	0.72	-0.43	0.02	-0.62	91

Notes: t-statistics (Newey–West) of b in $Y_{q+h} - \widehat{Y}^{SPF}_{q+h|q} = a + b\, m_q + \varepsilon_{q+h}$, where Y is the first-release target, $\widehat{Y}^{SPF}_{q+h|q}$ the median SPF forecast made in quarter q for quarter $q+h$ ($h = 0, 1$ for the current and next quarter), and m_q the filing-based measure averaged over the weeks of q up to the survey deadline; 1997–2019. ACT: pre-specified activity composite; ACT_Δ: its 13-week change; PRICE: price-pressure composite; LEAD, COIN: the GDP indices of Section 5. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table H.2: Predictive regressions with standard controls

Target	h	ACT	ACT_Δ	PRICE	LEAD	COIN	R^2, controls
Real GDP	1	-1.02	4.46***	-0.28	4.02***	-2.13**	0.85
	2	-3.36***	3.01***	-2.03**	0.92	-3.94***	0.83
	4	-1.81*	-0.49	-1.10	-1.87*	-2.11**	0.56
Industrial production	1	0.84	6.31***	-0.70	4.30***	-0.33	0.85
	2	0.35	4.98***	-1.40	2.86***	-0.45	0.82
	4	-1.51	0.27	-1.54	-0.90	-1.29	0.46
Business investment	1	2.76***	7.19***	0.02	6.90***	1.98**	0.84
	2	1.25	5.01***	-1.92*	5.45***	0.49	0.86
	4	-1.71*	0.96	-3.42***	-0.39	-1.40	0.59
Payrolls	1	4.10***	3.99***	0.19	4.47***	4.14***	0.92
	2	2.67***	3.15***	-1.77*	2.78***	2.97***	0.89
	4	-0.06	0.58	-3.13***	-0.59	0.59	0.72
Unemployment rate	1	-4.11***	-5.25***	0.32	-5.10***	-4.19***	0.90
	2	-2.52**	-4.45***	1.95*	-3.88***	-2.88***	0.88
	4	1.34	-1.14	3.46***	0.44	0.52	0.68
Core CPI	1	1.25	-2.92***	1.86*	-3.26***	1.09	0.63
	2	2.40**	-3.09***	3.54***	-3.63***	1.97**	0.50
	4	3.81***	-2.87***	5.53***	-1.55	2.67***	0.59
Headline CPI	1	-0.48	1.21	0.65	1.22	-1.17	0.45
	2	-0.68	1.17	0.70	-0.01	-1.27	0.33
	4	-0.80	1.20	0.94	1.03	-1.21	0.20

Notes: t-statistics (Newey–West) of the filing-based measure in regressions of the first-release target at $q+h$ on the measure at q, the latest published target, the term spread, the BAA spread, the CRSP value-weighted return, the VIX, Michigan sentiment, the CFNAI and the NFCI; 1997–2019. Last column: R^2 with the controls only. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table H.3: Recession probits

Measure	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
ACT	0.92→0.93	0.85→0.85	0.83→0.84	0.84→0.85	0.88→0.89*
ACT$_{\Delta}$	0.92→0.99***	0.85→0.92***	0.83→0.87*	0.84→0.85	0.88→0.88
PRICE	0.92→0.93*	0.85→0.89***	0.83→0.90***	0.84→0.91***	0.88→0.92***
LEAD	0.92→0.97***	0.85→0.88	0.83→0.83	0.84→0.84	0.88→0.89
COIN	0.92→0.93	0.85→0.85	0.83→0.84	0.84→0.86	0.88→0.90**

Notes: Probit of the NBER recession indicator in quarter $q + h$ on the term spread and the BAA spread, without and with the filing-based measure, 1997–2019 (in-sample). Cells: area under the ROC curve without → with the measure; stars: likelihood-ratio test of the measure, $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Local projections. Figure H.1 traces the response of the economy to a filing shock, the part of each measure not explained by current and four lagged values of GDP growth, core inflation, the change in unemployment, industrial production growth, stock returns, credit and term spreads, and consumer sentiment (Jordà, 2005). A positive shock to the activity composite is followed by lower output and industrial production, already within a year, and after five to eight quarters also by lower payrolls and higher unemployment (for the level of activity, −0.72% for GDP and −1.42% for industrial production after eight quarters, $p \approx 0.04$). Shocks to the price composite have similar delayed effects. One reading is that unusually upbeat reports, relative to what the macroeconomic data already show, signal optimism that is later disappointed; I do not interpret these responses causally.

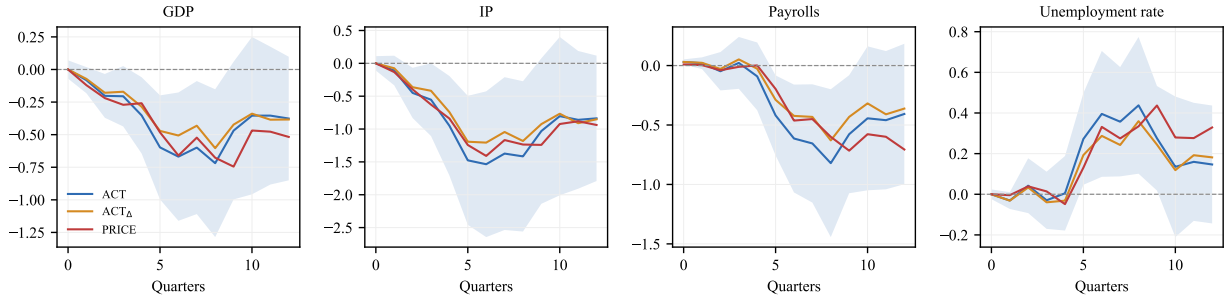


Figure H.1: Local-projection responses to filing-based measures

Notes: Local-projection responses (percent, or percentage points for the unemployment rate) to a one-standard-deviation shock to the activity composite (ACT, with 90% band), its 13-week change and the price composite, 1997–2019.

Recession probits. Following Estrella and Mishkin (1998), Table H.3 adds each measure to a probit of the NBER recession indicator on the term spread and the BAA spread. The change in the activity composite identifies current recessions almost perfectly: the area under the ROC curve rises from 0.92 to 0.99, and the measure is significant one quarter ahead as well. The price composite raises the probability of a recession one to four quarters ahead, with likelihood-ratio p-values below 0.01 at each of these horizons. With only two recessions in 1997–2019, these are in-sample results; I do not evaluate recession forecasts out of sample.